Reading the brain's personality: using machine learning to investigate the relationships between EEG and depressivity

Shenghuan Zhang

a thesis submitted for the degree of

Doctor of Philosophy

at the University of Otago, Dunedin,

New Zealand.

28 June 2019

Abstract

Electroencephalography (EEG) measures electrical signals on the scalp and can give information about processes near the surface of the brain (cortex). The goal of our research was to create models that predict depressivity (mapping to personality in general, not just sickness) and to find potential biomarkers in EEG data. First, to provide our models with cleaner EEG data, we designed a novel single-channel physiology-based eye blink artefact removal method and a mains power noise removal method. Then, we assessed two main machine learning model types (classification- and regression-based) with a total of eighteen sub-types to predict the depressivity of participants. The models were generated by combining four signal processing techniques with a) three classification techniques, and b) three regression techniques. The experimental results showed that both types of models perform well in depressivity prediction and one regression-based model (*Reg-FFT-LSBoost*) showed a significant depressivity prediction performance, especially for female group. More importantly, we found that a specific EEG frequency band (the *gamma* band) made major contributions to depressivity prediction. Apart from that, the *alpha* and *beta* band may make modest contributions. Specific locations (T7, T8, and C3) made major contributions to depressivity prediction. Frontal locations may also have some influence. We also found that the combination of both eye states' EEG data showed a better depressivity prediction ability. Compared to the eyes closed data, the EEG data obtained from the state of eyes open were more suitable for assessing depressivity. In brief, the outcomes of this research provided the possibilities for translating the EEG data for depressivity measure. Furthermore, there are possibilities to extend the research to apply to other mental disorders' prediction, such as anxiety.

Acknowledgements

First of all, I offer my profoundest gratitude to my supervisors: Professor Brendan McCane, Professor Neil McNaughton, and Associate Professor Zhiyi Huang for their invaluable guidance, advice, and motivation throughout my candidature. Without their help and guidance it is impossible for me to complete the work.

I must thank Dr. Phoebe Neo from the Department of Psychology for her support throughout my whole studies.

I would like to thank all members of Computer Science department, especially Dr. Alistair Knott who was kind in sharing his time, discussing and solving my problems; PhD students Wen Yang, Waqas Ahmad, Lahiru Ariyasinghe, and Aleksei Fedorov and many many others. Thanks for their support, encouragement and friendship to help me get through the tough time.

Last but not least, thanks to family at home in China for all the long-distance support while I am here in New Zealand; thanks to my best friend Jamie Qiu for her unconditional support to fill me with confidence and strength when I needed the most.

Contents

List of Tables

List of Figures

Glossary

AdaBoost Adaptive Boosting

ANN Artificial Neural Network

ANT Advanced Neuro Technology

ANOVA Analysis of Variance

ApEn Approximate Entropy

BDI Beck Depression Inventory

BSS Blind Source Separation

CART Classification and Regression Trees

CADP Computer-Aided Depressivity Prediction

CSP Common Spatial Pattern

CWT Continuous Wavelet Transform

D2 Correlation Dimension

DCA Dependent Component Analysis

DSM-5 Diagnostic and Statistical Manual of Mental Disorders (fifth edition)

DWT Discrete Wavelet Transform

EEG Electroencephalography

EMG Electromyography

ENN Elman Neural Network

EOG Electrooculogram

EPNN Enhanced Probabilistic Neural Network

FD Fractal Dimension

feature independent variable

FT Fourier Transform

FFT Fast Fourier Transform

Fuzzy SL Fuzzy Synchronization Likelihood

FSC Fuzzy Sugeno Classifier

FVC Feature Vector Construction Method

GMM Gaussian Mixture Model

HFD Higuchi's Fractal Dimension

ICA Independent Component Analysis

ICD-10 International Statistical Classification of Diseases and Related Health Problems - 10th Revision

KNN K Nearest Neighbors

LE Lyapunov Exponent

LSBoost Least Squares Boosting

Lasso Least Absolute Shrinkage and Selection Operator

LVQ learning vector quantization

MCA Morphological Component Analysis

ML-d Machine Learning Depressivity

MPCV Multiple predicting K-fold cross-validation

NBC Naive Bayes Classification

NIMH National Institute of Mental Health

NN Neural Network

PA Participant Assignment

PCA Principal Component Analysis

Ph Bispectral Phase Entropy

PID-5 Personality Inventory for DSM-5

PID5-d Pid-5 Depressivity

RBFNN Radial Basis Function Neural Network

RBPNN Radial Basis Probabilistic Neural Network

REN Renyi Entropy

RF Random Forest

RQA Recurrence Quantification Analysis

RPs Recurrence Plots

SP Sample Points

SNR Signal to Noise Ratio

STFT Short Time Fourier Transform

STFT-BW Stft at Bandwidth of Total Spectrum

SVD Singular Value Decomposition

TEBR Template Eye Blink Removal Method

WPD Wavelet Packet Decomposition

Chapter 1

Introduction

The brain controls most of the activities of the body and is also responsible for cognition, emotion, and memory (Purves, Augustine, Fitzpatrick, Hall, LaMantia, McNamara, and Williams, 2004; Siuly, 2012). This power is a result of the brain's complexity, which also makes it difficult to analyse.

Electroencephalography (EEG) measures electrical signals on the scalp and can give information about brain processes near the surface of the brain (cortex). In recent times, EEG as a bio-signal has been frequently used to characterise different emotional states (Ackermann, Kohlschein, Bitsch, Wehrle, and Jeschke, 2016), physiological states, and brain-related diseases (Mustafa, Taib, Murat, and Sulaiman, 2012) such as dementia (Jeong, 2002), schizophrenia (Elbert, Lutzenberger, Rockstroh, Berg, and Cohen, 1992; Li, Tong, Liu, Gai, Wang, Wang, Qiu, and Zhu, 2008) and obsessive-compulsive disorder (Aydin, Arica, Ergul, and Tan, 2015; Akdemir Akar, Kara, Agambayev, and Bilgic, 2015). Therefore, the EEG is potentially an important tool in the diagnosis of emotional states, brain-related diseases and other physiological states.

In this research, we assess and predict depressivity based on participants' EEG data. This is an application of computer technology to measurement of personality. In the research, we propose two main types of models: a classification type of model based on the assumption that occurrence of some specific EEG patterns which indicate 'High depressivity' increases in number with a higher depressivity; and a regression type of model based on the assumption that EEG patterns vary with different depressivity.

These proposed methods were used to predict an individual's depressivity and identify the biomarkers from the EEG signal.

1.1 Overview and Motivation of the Study

Electroencephalography (EEG) records or traces electrical signals generated by the brain. It shows summed potentials of the physiological activity of cells in the cerebral cortex. Cortex is the highest processing level in the human brain. A detailed discussion of EEG is provided in Section 2.3 of Chapter 2.

We chose depression as our starting problem because it is the most problematic mental disorder. Depression impacts many aspects of an individual, including growth, development, and family relationships (Bhatia and Bhatia, 2007). Worldwide prevalence rates of depression are between 2.2 to 10.4% (Nabbe, Le Reste, Guillou-Landreat, Munoz Perez, Argyriadou, Claveria, Fernandez San Martin, Czachowski, Lingner, Lygidakis, Sowinska, Chiron, Derriennic, Le Prielec, Le Floch, Montier, Van Marwijk, and Van Royen, 2017), and the economic burden of depression was estimated to be approximately $US210 billion dollars in the United States in 2010 (Greenberg, Fournier, Sisitsky, Pike, and Kessler, 2015). If those who suffer from depression are not diagnosed and treated in time, severe consequences can easily be caused; and it can even lead to suicide (Kessler, DuPont, Berglund, and Wittchen, 1999).

Depression is not well characterised as a biological entity. Therefore, in Section 2.2, depression and in particular depressivity as defined by the Personality Inventory for DSM-5 (PID-5) will be discussed.

Critically, the biological causes of depression are largely unknown (Fröhlich, 2016). Therefore, clinical diagnosis of depression is currently based on symptoms, equivalent to fever, rather than its underlying causes, which is equivalent to the measles virus (Kupfer, Frank, and Phillips, 2012). Consequently, the National Institute of Mental Health (NIMH) has called for the development of biomarkers to enhance the diagnosis of mental health disorders (Insel, Cuthbert, Garvey, Heinssen, Pine, Quinn, Sanislow, and Wang, 2010; Insel, 2014). Hence, in this research we propose to use computer technology, in particular machine learning methods, to design several depressivity prediction models to predict depressivity and find biomarkers of depressivity disorder in EEG signals. Ultimately, these biomarkers could redefine our ideas of what depression actually is.

1.2 Problem Statement: How Can We Predict Depressivity from EEG Data?

Due to late detection and misdiagnosis, the prevalence of depression is very high, even though it is ultimately treatable with effective treatments in up to three-quarters of cases (Acharya, Sudarshan, Adeli, Santhosh, Koh, and Adeli, 2015). Traditionally, specialists diagnose depression based on patients' symptoms. However, sometimes the symptoms of depression (e.g. poor sleep) can be confused with other mental conditions. Furthermore, depression is a label that covers a range of disorders, with a variety of different types, and is not well defined (Bachmann, Kalev, Suhhova, Lass, and Hinrikus, 2015). Thus, to find objective biomarkers for the medical diagnosis of depression could be a more appropriate approach.

The EEG signal contains large amounts of data about many areas of the brain and cannot be easily manipulated by the subject. Therefore, it offers a great potential for unbiased depression recognition. In recent years, many EEG studies have focused on detecting clinical depression (Hosseinifard, Moradi, and Rostami, 2013; Li and Fan, 2006; Puthankattil and Joseph, 2012b). EEG studies on depression have shown that EEG data can be used to effectively distinguish depressive participants and healthy controls (e.g.(Bachmann, Lass, and Hinrikus, 2017; Mohammadi, Al-Azab, Raahemi, Richards, Jaworska, Smith, de la Salle, Blier, and Knott, 2015; Hosseinifard *et al.*, 2013)). Some papers reported that the accuracy of classifying depressive participants and healthy controls was over 90% (Faust, Ang, Puthankattil, and Joseph, 2014; Bachmann *et al.*, 2017).

However, in order to prevent individuals from reaching a severe and irreversible state that is clinical depression, the recognition of depression in its early stage is more important than detection of clinical depression (Acharya *et al.*, 2015). Furthermore, there is no clear cut-off between healthy people and depressed patients. Therefore, instead of detecting depression, it is better to predict the depressivity of individuals. In this research, we used machine learning methods to design several models to predict depressivity.

1.3 Prior Knowledge Required for This Research

This thesis contributes to the knowledge of predicting depressivity using EEG data.

The interdisciplinarity of this research requires knowledge from different fields:

- Psychology: Knowledge of the psychology field is needed. It helps to design the experiments, pre-analyse the EEG data and group the data sets. For example, understanding gender difference and eye state difference is beneficial to establish the prediction models.

- Signal processing technology: Knowledge of signal processing is very important. The interconnection of numerous neurons makes EEG signals highly nonstationary and highly unstable, so they are difficult to characterise (Subha, Joseph, Acharya, and Lim, 2010). Most EEG signals are collected from the scalp and contaminated by many sources, so they have a very low Signal to Noise Ratio (SNR). Consequently, signal processing technology can be used to understand the effect of millions of neurons and their correlation to physiological activities, and increase SNR to collect features which are invisible to the naked eye (Luo, Ibrahim, Ismail, and Xu, 2013; Hu, Wang, and Ji, 2013). There are various signal processing technologies, and we need to explore them and choose the most suitable technologies for EEG signal analysis.

- Machine learning: There are several kinds of machine learning methods fit for different applications. It is important to find the best ones to analyse the features and establish effective models.

1.4 Contributions of This Thesis

The contributions of this thesis include:

1. Designing two novel EEG noise removal methods: a single-channel physiology-based eye blink artefact removal method as reported in Chapter 4 and in (Zhang, McIntosh, Shadli, Neo, Huang, and McNaughton, 2017); a mains power noise removal method which is reported in Section 3.4.2 of Chapter 3.

2. Proposing and evaluating 18 prediction models with novel combinations of machine learning methods and feature extraction techniques for EEG. 12 of these were classification-based models and the rest were regression-based methods. One regression-based model (*Reg-FFT-LSBoost*) showed a significant depressivity prediction performance for the female group. These are reported in Chapters 7 and 8. The comparison of two kinds of prediction models (classification- and regression-based) showed that the performance of the regression-based models was better than the classification-based models.

3. Discovering that the *gamma* band of frequencies is potentially very important for predicting depressivity. Apart from that, the *alpha* and *beta* band may also make modest contributions to depressivity prediction which are shown in Section 7.7 and 8.5; identifying that three specific EEG locations (T7, T8 and C3) are potentially very important for predicting depressivity. Frontal locations (especially Fp2) may also have some influence for depressivity prediction, as shown in Sections 7.7 and 8.5. The specific EEG frequency bands and locations are potential biomarkers.

4. The overall results obtained from experiments also imply that the features obtained in the frequency domain were worse than the features obtained by the original time domain signal. Although previous related research almost all focused on frequency domain features, we still need to pay more attention to the methods that can derive the spatial distribution information from time domain signal in the future work.

1.5 Structure

This thesis is structured into 9 chapters. In Chapter 2, the research background and related work is presented. Chapters 3 and 4 give an overview of depressivity prediction models and the data preprocessing details. Chapters 5 and 6 contain detailed descriptions of the important parts of the models: feature extraction (signal processing technologies) and prediction (machine learning technologies). Chapter 7 and 8 detailed discussion of two kinds of depressivity prediction models' establishment, evaluation and comparison. Chapter 9 summarises the results, discusses the findings and concludes the thesis.

Chapter 2

Background and Related Work

This chapter presents an overview of the background of this research, including the concepts of EEG, depression and depressivity. This chapter also surveys previous work which is relevant to this thesis.

2.1 Introduction

EEG has been shown to reflect a variety of mental disorders. The reflection of mental disorders in EEG was observed from a long time ago. In 1946, Conn (1946) already found two special EEG patterns in a group of individuals who showed anxiety coupled with excitement, depression or irritability. The first pattern observed was decreasing alpha wave (the concept of alpha wave and the following beta wave are introduced in Section 2.3.2.2). The second pattern was rhythmic high frequency oscillations (beta waves) ranging from 10 to 50 microvolts. These discoveries were the initial explorations in this area, and these results are still instructive.

In recent years, with the development of computer technology, an increasing number of methods are being used to analyse EEG and explore its relationship with mental disorders. Our work focuses on one mental disorder – depression. In this chapter, related works are introduced in detail.

2.2 Depression

Depression is a label that covers a range of disorders with a variety of different types (Bachmann *et al.*, 2015), such as persistent depressive disorder (also called dysthymia), postpartum depression, psychotic depression and seasonal affective disorder (National

Institute of Mental Health, 2018a). These kinds of depression have different symptoms and develop in different environments. In general, depression can lead to a wide range of problems in people's mental state and physical health. They experience feelings of sadness, loss of interest in activities once liked, a loss of or increased sleep and appetite, rendering a person unable to function as normal (American Psychiatric Association, 2017). Depression also increases the risk of suicide (Kessler *et al.*, 1999).

According to a survey by WHO, in 2017, there were over 300 million people living with depression (World Health Organization, 2017), and one in six people (16.6%) will experience depression once in their lifetime (American Psychiatric Association, 2017). In America, the prevalence among elderly patients (age: 50+ years) ranged from 13.2% to 16.5% (Volkert, Schulz, Härter, Wlodarczyk, and Andreas, 2013). Furthermore, there was an 18% increase in the number of depressed people between 2005 and 2015 (World Health Organization, 2017), and the economic burden of depression was estimated to be approximately $US210 billion dollars in the United States in 2010 (Greenberg *et al.*, 2015).

Depression can occur at any stage in life, but generally, it begins from early teens to mid-20s (adolescents) (National Institute of Mental Health, 2018b).

Among young people (ages 14–25), the prevalence of depression in females is twice as much as for males (Albert, 2015). Many causes contribute to this situation, some socioeconomic factors such as abuse, education, diet and income, and biological contributors, such as hormone differences in brain circuitry (Albert, 2015; Gillies and McArthur, 2010).

Nevertheless, depression can be treated (Acharya *et al.*, 2015) and it is one of the most treatable mental disorders (American Psychiatric Association, 2017). 80% to 90% of depressed patients have a positive response to treatment. After receiving treatment, almost all of the patients have a remission of symptoms (American Psychiatric Association, 2017). However, the effects of treatments depend on the treatment's starting time. If the treatment starts early, it will be more effective (Hosseinifard *et al.*, 2013; Mohammadi *et al.*, 2015).

Therefore, if diagnosed in a timely fashion, depressed patients will have a higher chance of recovery.

2.2.1 Difficulties in Depression Diagnosis

Recent studies show that depression is caused by several factors including biochemistry, genetics, biological, environmental or psychological factors (American Psychiatric As-

sociation, 2017; National Institute of Mental Health, 2018a). However, these factors, especially the biological causes of depression, are largely unknown (Fröhlich, 2016). Therefore, clinical diagnosis is based on symptoms rather than the underlying causes.

To diagnose depression, there needs to be a definite change of mood which has lasted at least two weeks. This change of mood is distinguished by symptoms such as feelings of sadness and hopelessness, trouble with sleep and appetite; constant crying; suicidal or self-harming thoughts; or changes in the speed of speech or action (Belmaker and Agam, 2008).

However, the method of this kind of diagnosis relies on subjective judgments. Thus, the result will be affected by many factors, including incorrect descriptions of symptoms by patients, misunderstanding by specialists or confusing these symptoms with other mental states or diseases. For example, depression will sometimes be confused with grief (American Psychiatric Association, 2017) because both depression and grief will cause people to feel sad and lose interest in activities they once liked. The confusion is caused by a symptoms-based clinical diagnosis. Some medical conditions, drug or alcohol influences and physical illnesses will also create the same symptoms as depression (Healthdirect Australia, 2017).

Based on the above information, it is clear that the diagnosis of depression relies largely on the information provided by the patients. The patients need to describe their feelings and emotional changes to their doctors. If the patient fails to express themselves clearly, it will lead to misjudgments.

Consequently, the NIMH has called for the development of biomarkers to enhance the diagnosis of mental health disorders (Goodman, 2009). The term 'biomarker' has a variety of meanings in different application fields (Schmidt, Shelton, and Duman, 2011). In this research, it refers to the specific patterns in the EEG which can predict PID-5 depressivity scores. These patterns may reflect or represent neuron changes that could be linked to mental disorder.

2.2.2 Depressivity

Because of the different types of symptoms and patients suffering at different severities, depression is also classified from mild, moderate to severe. It can be divided at a finer-grain level by using scores of what we call depressivity.

Because depression is so extensive and not well defined, in our research, we need a metric to measure the depressivity of our participants and use this metric as the base of our research.

Some metrics are used to measure depressivity, and the most popular ones are the Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, and Erbaugh, 1961; Schaffer, Davidson, and Saron, 1983) and Personality Inventory for DSM-5 (PID-5) (Krueger and Markon, 2014; American Psychiatric Association *et al.*, 2016).

The BDI is a clinical scale, so a bottom score on the BDI means "not thought to be sick". Higher scores suggest illness and are not set up with particular psychometric properties.

PID-5 is very similar to BDI in terms of critical constructs, but it is a population scale. Low values mean far from clinical. Medium is normal for the middle of the population and high is getting clinical. A score above a particularly high threshold can be taken as clinical. In other words, the PID-5 depressivity scale is designed to give high values in those with disorder but also a range of values that represents a normal personality scale and covers the general population. So it is a personality scale that can include clinical populations but is not a clinical test used for diagnosis. Therefore, a low BDI (bottom end of clinical) is similar to a high PID. In both cases, the scores close to the top are seriously clinical.

In this research, we chose the PID-5 depressivity score as the metric because it is wider (so it gives more variance for machine learning methods to work with), covers the normal range (maps to personality, not just sickness) and is more modern and psychometrically validated.

The details of the PID-5 will be introduced in the next section.

2.2.2.1 The PID-5 depressivity score

The Diagnostic and Statistical Manual of Mental Disorders (DSM) is a standard classification of mental disorders published by the American Psychiatric Association with its fifth edition appearing in May 2013 (American Psychiatric Association *et al.*, 2013). This edition is called DSM-5. In addition to the standard symptom classification schemes used for diagnosis, DSM-5 provides a new personality assessment tool, the Personality Inventory for DSM-5 (PID-5) (Fossati, Krueger, Markon, Borroni, and Maffei, 2013; Krueger, Derringer, Markon, Watson, and Skodol, 2012).

The PID-5 consists of a number of scales with scores calculated from a 220-item self-report questionnaire of items with a 4-point response scale from "very false or often false" to "very true or often true" (Zabelina, Condon, and Beeman, 2014; American Psychiatric Association *et al.*, 2016). PID-5 is intended to assess six of the ten DSM defined personality disorders: Schizotypal, Antisocial, Borderline, Narcissistic, Avoidant,

and Obsessive Compulsive (Kelsey and G.Fischer, 2015). However, the PID-5 scales also includes measures of traits linked to anxiety and depression; and their range includes the healthy as well as the clinical populations. There are 25 scales of this type in this questionnaire: Anhedonia, Anxiousness, Depressivity, Emotional lability, Hostility, Perseveration, Rigid Perfectionism, Separation insecurity, Submissiveness, Suspiciousness, Withdrawal, Attention Seeking, Callousness, Deceitfulness, Grandiosity, Manipulativeness, Intimacy Avoidance, Restricted affectivity, Distractibility, Eccentricity, Perceptual Dysregulation, Risk Taking, Unusual Beliefs & Experiences, Impulsivity and Irresponsibility (Krueger and Markon, 2014).

The main focus of this thesis is depression as a disorder or group of disorders. Therefore, the work focuses on only one scale of the PID-5: depressivity. Our approach takes the view that a mental disorder is an extreme of a normal personality variation. The PID-5 depressivity score is used to measure such a personality factor, and at least some patients with clinical depression would score very high on the scale (Fossati *et al.*, 2013; Markon, Quilty, Bagby, and Krueger, 2013). An additional advantage of the PID-5 scale for our work is that it is constructed to be simple and essentially unidimensional.

2.3 Human Brain

In our research, we first collect EEG from the brain to analyse the relationship between EEG and depressivity. Below is an introduction to the brain.

2.3.1 Structure

The human brain can be morphologically divided into the following primary structures: cerebrum, cerebellum, brainstem and diencephalon (Health Jade Team, 2018). The cerebrum as the largest part of the brain is our research objective. In this thesis, the term 'brain' will be used to refer mainly to the cerebrum. The brain is divided into two hemispheres, and these two symmetric parts are connected by the corpus callosum (Witelson, 1985), which is made up of nerve fibres. Each hemisphere is divided into frontal, parietal, temporal and occipital lobes (shown in Figure 2.1).

- Frontal Lobe: The frontal lobe is located in the anterior half of the brain (anterior to the parietal lobe, and anterior and dorsal to the temporal) (Stuss and Benson, 1986). Compared with other lobes, the frontal lobe is the biggest part

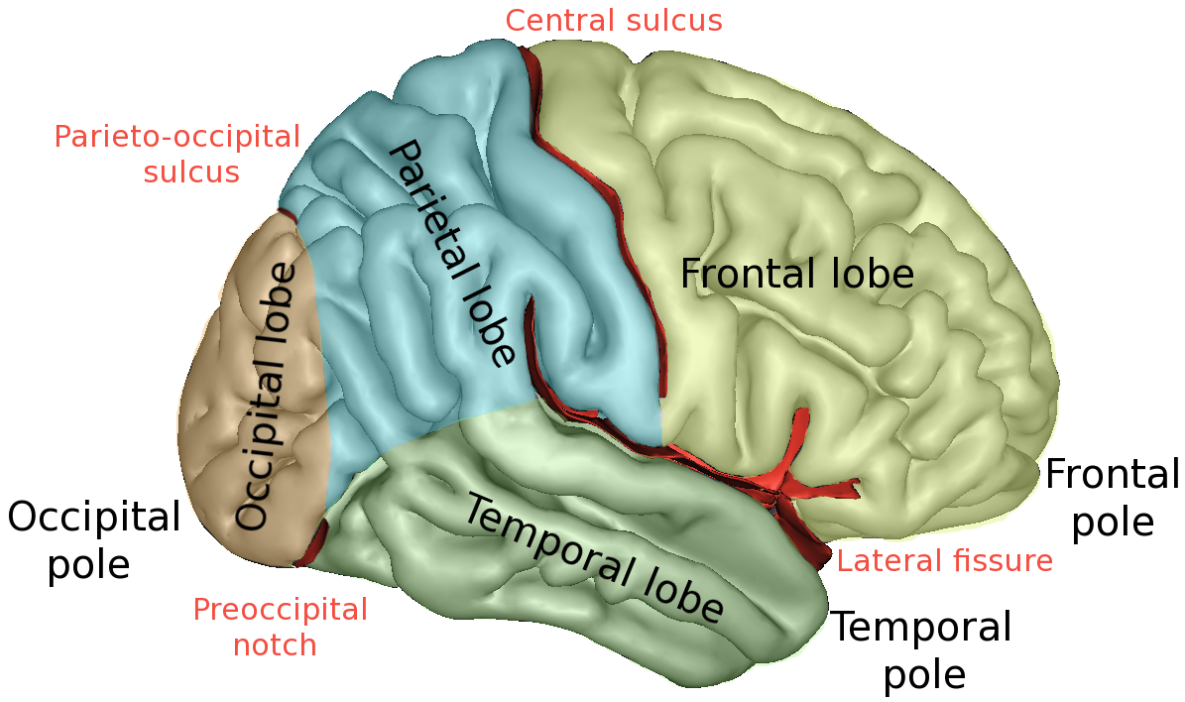

Figure 2.1: Brain lobes: frontal lobes, parietal lobes, temporal lobes and occipital lobes. Figure is download from `https://commons.wikimedia.org/wiki/File:LobesCaptsLateral.png`

of the human brain (Kimberg and Farah, 1993) and is associated with functions such as the formation of human language, language expression (Voets, Adcock, Flitney, Behrens, Hart, Stacey, Carpenter, and Matthews, 2005), emotional expression (Stuss, Gow, and Hetherington, 1992), autonomous consciousness (Dietrich, 2004), and voluntary muscle control (Teitti, Määttä, Säisänen, Könönen, Vanninen, Hannula, Mervaala, and Karhu, 2008; Bianchi, 1895). The frontal lobe plays a key role in integrating the brain's non-task long-term memory (Jetter, Poser, Freeman Jr, and Markowitsch, 1986).

- Parietal Lobe: It is located behind the frontal lobe. This area deals with all kinds of sensory information (such as pain and touch), and is also related to language (Cohen, Dehaene, Chochon, Lehericy, and Naccache, 2000), memory (Wagner, Shannon, Kahn, and Buckner, 2005), and other functions.

- Temporal Lobe: It is the site of the primary and secondary auditory cortex, and the centre for processing auditory information (Kimura, 1961). The temporal lobe is also important for understanding language (Haglund, Berger, Shamseldin,

Lettich, and Ojemann, 1994), and damage to this part of the brain can lead to aphasia. The hippocampus of the medial temporal lobe plays an important role in forming long-term memories (Squire and Zola-Morgan, 1991). The temporal lobe is also involved in object and face recognition through vision (Rolls, 2001).

- Occipital Lobe: Its main known functions are various types of processing of visual information (Bhat, Acharya, Adeli, Bairy, and Adeli, 2014).

2.3.2 The EEG

The Electroencephalogram (EEG) is the record or tracing of electrical signals generated by the brain. It shows summed potentials of physiological activity of cells in the cerebral cortex. EEG signals are analogue signals before we store and use them. EEG signals are then converted to digital values for data processing. There are two kinds of EEG from a processing point of view: spontaneous EEG and evoked EEG. Evoked EEG is a response to a stimulus, and it is time-locked to the stimulus. Spontaneous EEG is the brain's spontaneous electronic signal without any experimenter-controlled stimulus, often recorded in the resting state. In this research, we analysed participants' spontaneous resting EEG to detect their mental disorder state (depressivity).

2.3.2.1 10-20 System

EEG signals can be collected from inside the head or from the surface of the head (scalp). The signals collected from inside the head are stronger, but implantation of the electrodes involves invasive tissue damage. Therefore, it is only undertaken with clinical patients for therapeutic reasons. Therefore, for research purposes, the signals are often recorded, non-invasively, from the scalp. The positions of scalp electrodes generally use the international 10-20 system (Figure 2.2).

The international 10-20 system is a method to define the scalp electrode position reliably across people in relation to brain structures and is internationally accepted. The locations of electrodes in this system map to the specific underlying areas of the brain. It is called the 10-20 system because the distances between adjacent electrodes are 10% or 20% of the total front-back (from nasion to inion), or right-left (right preauricular point to left preauricular point) distance of the skull or around the skull circumference. To receive a more detailed, high-resolution EEG recording, the system recording will shrink to be 5% distance.

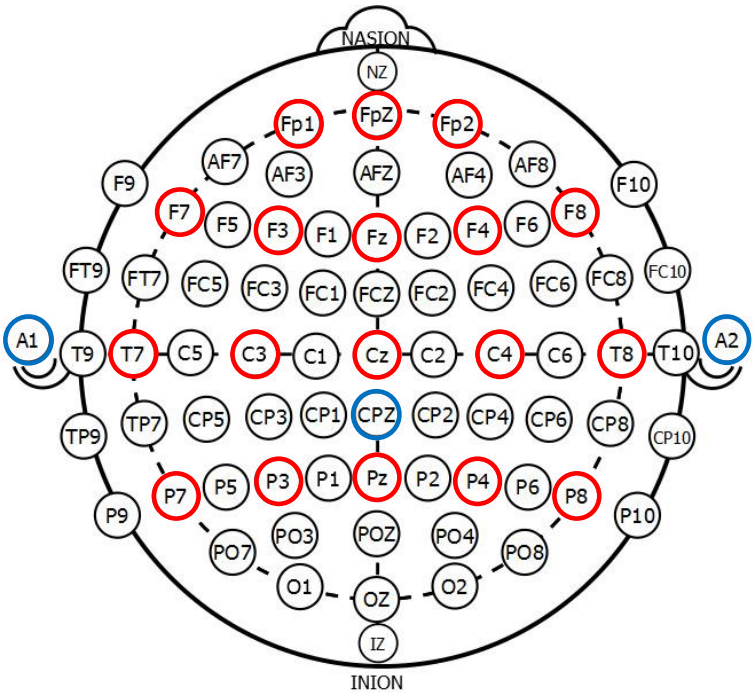

Figure 2.2: The extended international 10-20 system position of scalp electrodes. Electrodes are placed at 5%, 10% and 20% spacings relative to standard skull measurements (e.g. nasion-inion). Figure is CC0 from `https://commons.wikimedia.org/wiki/File:International_10-20_system_for_EEG-MCN.svg`. Abbreviations: A = Auxiliary (Ear lobe, shown, or mastoid, in our experiments), C = central, P = parietal, F = frontal, Fp = frontal polar, O = occipital, T=temporal. The sites used in this thesis are marked by red circles, and reference channels are marked by blue circles.

Each electrode of the 10-20 system has a label that indicates the area that site locates. The letter in each electrode label shows the lobes that the electrode belongs to. The abbreviations refer to: A =Auxiliary, C=Central, Fp=Frontal Polar, F=Frontal, O=Occipital, P=Parietal or T=Temporal. Note, there is no Auxiliary, Central or Frontal Polar lobe. The Auxiliary refers to the Ear lobe or mastoid in our experiments. Central and Frontal Polar identify the location of the central brain and the frontal polar portion of the frontal lobe.

The numbers on the electrode label indicate which part of the hemispheres that electrode locates. The numbers 2,4,6,8 refer to the right hemisphere, and the numbers 1,3,5,7 refer to those electrodes located on the left. Z (zero) refers to the midline which separates the two hemispheres.

Figure 2.2 also shows some electrodes' names AF, FC, FT, CP, TP and PO. These extra electrodes are placed for recording more detailed EEG. These electrodes are added between the original Fp, F, C, T, P and O sites (See Table 2.1).

Electrodes	Location
AF	between Fp and F
FC	between F and C
FT	between F and T
CP	between C and P
TP	between T and P
PO	between P and O

Table 2.1: 10-20 system's extra electrodes

2.3.2.2 Frequency band

EEG signals usually show rhythms at different frequencies that are associated with distinct behavioural states (Buzsáki and Draguhn, 2004; Buzsáki, 2006). Conventionally, EEG signals are divided into five different frequency bands. They are α, β, γ, δ and θ (Figure 2.3).

- δ `band`: 1~3.5 Hz. This occurs very clearly in sleep, and in particular is the key feature of deep 'slow wave' sleep.

- θ `band`: 3.5~8 Hz. Theta rhythm is observed to link to the integration of attention and sensorimotor (Ekstrom, Caplan, Ho, Shattuck, Fried, and Ka-

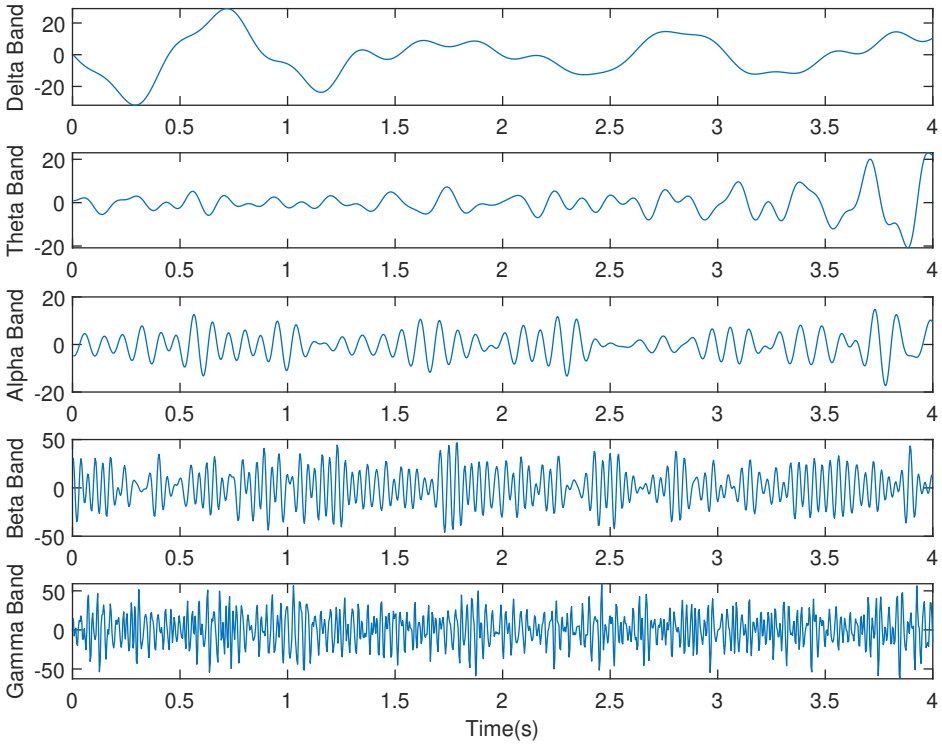

Figure 2.3: EEG signals are divided into five different kinds of waves in frequency band.

hana, 2005), memory load (Lega, Jacobs, and Kahana, 2012; Sauseng, Hoppe, Klimesch, Gerloff, and Hummel, 2007) and cognitive demand (Sauseng *et al.*, 2007).

- α `band`: 8~13 Hz. Alpha rhythm can occur in clear visible bursts when the shape of the wave is a sinusoidal wave. α wave bursts are the most obvious rhythm in the EEG data. α wave is always seen in normal relaxed states (for example, if eyes are closed (Subhash Chandran, Mishra, Shirhatti, and Ray, 2016)).

- β `band`: 13~30 Hz. Generally, this appears when the brain is aroused.

- γ `band`: 30~80 Hz. This is one of the more important bands for psychologists. It is linked to memory, attention, and cognition (Buzsáki, 2006; Fries, 2009).

The EEG shows features which are typical of bioelectrical signals:

- The amplitude of scalp EEG signals is small. It usually ranges from 10 to 100μv and up to 200μv at the scalp, and is only higher in abnormal states such as

epilepsy. As a result, many sources can generate artefacts and influence the weak EEG. There are two basic sources of EEG artefacts (Fatourechi, Bashashati, Ward, and Birch, 2007): one is non-physiological sources, such as mains power oscillations interference, and changes in electrode impedances; the other one is physiological sources, such as eye blinks, eye movements, muscle movements.

- The EEG is a non-stationary signal since it is continuously affected by the environment and variation in physiological factors.

2.4 Related Work

Albert (2015) pointed that the incidence of depression is high, especially in women. If people suffer from depression, but are not diagnosed and treated in time, it will easily cause severe consequences and can even lead to suicide (Kessler *et al.*, 1999). However, depression is difficult to diagnose, and its symptoms are easily confused with other diseases and mental disorders. Electroencephalography (EEG) signals are a potentially rich source of biomarkers, and so they may be a useful tool for making an objective diagnosis of depression.

Studies based on EEG signals have increased significantly, and the researchers who study EEG come from many disciplines, including computer science, psychology, engineering, biomedical engineering and so on (Al-Nafjan, Hosny, Al-Ohali, and Al-Wabil, 2017). Because EEG research is interdisciplinary, EEG systems have different application purposes and research methods.

Our research interests mainly focus on EEG-based depressivity detection, but we can still borrow analytical methods from other applications. Therefore, this chapter will study different kinds of EEG-related literature to provide a research foundation for our depressivity prediction.

2.4.1 Literature Category

Because articles on EEG research spread into many disciplines, different scientific fields have different research focuses and methods. It is therefore difficult to make a comparative analysis of articles. Therefore, this section aims to classify and summarise the EEG-based research related to our study. Depression is a mental disorder, and it has some common characteristics with other mental disorders. Therefore, the literature about EEG studies on other mental disorders, emotional states, psychology and

psychiatry were also included. EEG as an electronic signal can be analysed using some analysis methods. Consequently, there is another category named 'EEG data analysis methods' includes three application purposes: EEG artefact removal, EEG feature extraction and EEG feature analysis. Figure 2.4 shows these two main categories (general EEG studies of different physiological states and different EEG data analysis methods) and subcategories.

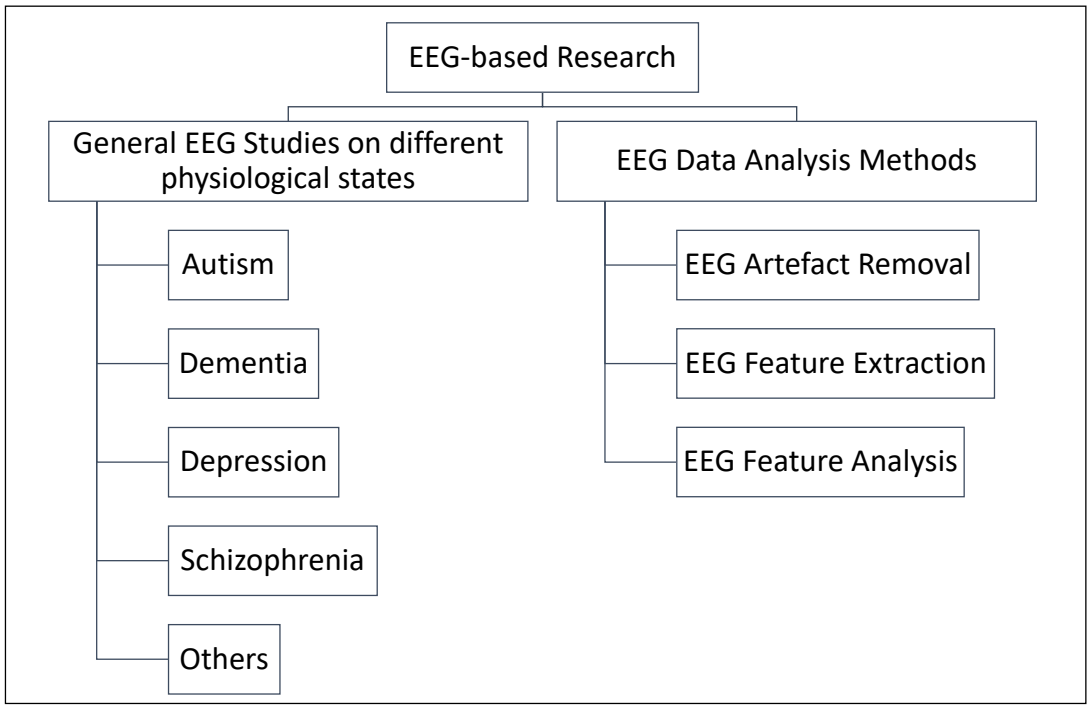

Figure 2.4: This figure lists the categories of EEG-base research selected to study.

2.4.2 General EEG Studies of Different Physiological Status

Electroencephalography (EEG) signals are a potentially rich source of biomarkers and have been previously used to characterise different neurological disorders (with detectable brain abnormalities) such as autism (Wang, Barstein, Ethridge, Mosconi, Takarae, and Sweeney, 2013), dementia (Jeong, 2002), schizophrenia (Elbert *et al.*, 1992; Li *et al.*, 2008) and obsessive-compulsive disorder (Aydin *et al.*, 2015). Depressivity as a psychiatric disorder is also expected to be predicted by EEG.

Autism In research on autism, researchers use the power alterations of EEG to assess autism. Wang *et al.* (2013) show that when compared to controls, the autism patients

present a U-shaped profile of EEG power alterations, with excessive power in both low-frequency and high-frequency bands. Their study also shows hemispheric asymmetry with enhanced power in the left hemisphere of the brain relative to the right.

Dementia Dementia refers to memory loss and other cognitive abilities that severely affect daily life. Alzheimer's disease accounts for 60 percent to 80 percent of dementia cases (Alzheimer's Association, 2018), and it is the most common dementia. There is no cure for Alzheimer's, but an early and accurate detection of Alzheimer's is still beneficial for the disease's control. EEG as a non-invasive and simply obtained signal and is also used to detect Alzheimer's. Typical EEG abnormalities in Alzheimer's are an increase in theta and delta activity, and a decrease in alpha and beta activity (Jeong, 2002). Jeong (2002) summarised some results of using non-linear dynamical changes of the EEG to characterise Alzheimer's. Correlation dimension (D2) as an important measure of non-linear dynamical analysis is generally considered to reflect the complexity of cortical dynamics (Röschke and Aldenhoff, 1991). The results of (Woyshville and Calabrese, 1994; Besthorn, Sattel, Geiger-Kabisch, Zerfass, and Förstl, 1995; Jelles, Strijers, Hooijer, Jonker, Stam, and Jonkman, 1999) all show that the D2 values are lower in Alzheimer's patients than those in controls.

Schizophrenia An article by Gordon, Palmer, and Cooper (2010) shows that compared to controls, people with schizophrenia can be characterised by having greater left lateralized alpha power. Schizophrenic patients also show increased activity in the low frequency band (1-7 Hz), a scalp distribution of dimensional complexity and Hjorth complexity (Elbert *et al.*, 1992). Li *et al.* (2008) and Fernández, López-Ibor, Turrero, Santos, Morón, Hornero, Gómez, Méndez, Ortiz, and López-Ibor (2011) proposed to calculate Lempel–Ziv complexity (LZC) scores to distinguish schizophrenic patients from controls. The results show that schizophrenic patients have higher complexity values than controls.

Obsessive-Compulsive Disorder Research by Kuskowski, Malone, Kim, Dysken, Okaya, and Christensen (1993) calculated the power of different bandwidth and location, and it showed that compared to the control group, obsessive-compulsive disorder patients show lower log absolute power in the delta and beta bands at frontal and right-hemisphere locations. Aydin *et al.* (2015) applied embedding entropy measurements, mean coherence and mutual information methods on EEG data and found that

in Obsessive-Compulsive Disorder patients, there are obvious EEG differences in the prefrontal areas when compared to normal people.

Depression The studies above emphasised that EEG as a biosignal has been successfully used in characterising several physiological states, and some biomarkers in EEG have been detected. During the past, EEG studies on depression have also shown that EEG data can be effectively used to distinguish depressed patients from healthy controls.

Hemispheric asymmetry is abnormal in depression but not all research support this point of view. Some early studies showed there is hemispheric asymmetry phenomenon of EEG power in subjects with depression (e.g., Henriques and Davidson, 1990; Baehr, Rosenfeld, Baehr, and Earnest, 1998; Gotlib, 1998; Allen, Urry, Hitt, and Coan, 2004; Gordon *et al.*, 2010). Different age groups show different EEG behaviours (Acharya, Sree, Swapna, Martis, and Suri, 2013) and depression symptoms, therefore, Deslandes, de Moraes, Pompeu, Ribeiro, Cagy, Capitão, Alves, Piedade, and Laks (2008) particularly selected elderly participants for the study. In this research, the result also showed the EEG signal asymmetry's existence. The research of Dawson, Frey, Panagiotides, Osterling, and Hessl (1997) showed that the children of depressed mothers also demonstrated relatively less left frontal activity than right. Therefore, the hemispheric asymmetry as a solid index has been used to detect depression for decades.

However, some research shows that hemispheric asymmetry does not have good diagnostic results (Segrave, Cooper, Thomson, Croft, Sheppard, and Fitzgerald, 2011; Debener, Beauducel, Nessler, Brocke, Heilemann, and Kayser, 2000). The simple asymmetry scores cannot effectively discriminate depressed patients and normal controls (Henriques and Davidson, 1991). Depressed patients without an anxiety disorder do not differ significantly from normal controls in their anterior alpha asymmetry (Bruder, Fong, Tenke, Leite, Towey, Stewart, McGrath, and Quitkin, 1997). There are no significant differences in frontal activation between depressed and non-depressed participants with either college students having high Beck Depression Inventory scores or with individuals diagnosed with DSM-III-R depression (Reid, Duke, and Allen, 1998). The results of (Debener *et al.*, 2000) showed the alterations of anterior brain activation in depressed patients were unstable.

Furthermore, EEG asymmetry is not specific for depressivity. The EEG asymmetry was also detected in other mental disorder or psychopathology states (Blackhart, Minnix, and Kline, 2006; Gordon *et al.*, 2010; Thibodeau, Jorgensen, and Kim, 2006).

Therefore, even though in some research the EEG asymmetry was announced as a mark for a depression risk, in our research we do not take this approach to get the feature from EEG. Because EEG asymmetry is not specific to depression as well as not always being abnormal in depression.

2.4.3 EEG Data Analysis Methods

There are many data analysis methods used to analyse EEG data. They are mainly applied to EEG artefact removal, EEG feature extraction and analysis.

2.4.3.1 Artefact removal

The various waveforms of EEG convey valuable information. However, artefacts may affect signals that we are interested in. The artefacts are either non-physiological sources or physiological sources. Non-physiological sources include the changes in impedance between the electrodes and the scalp, electromagnetic interference signal generated by the mains power and the interference source from the environment, while the physiological sources refer to the signals generated by some organs and tissues of the human body in the process of obtaining EEG, such as eye blinks, eye movement, cardiac activity and muscle movement (Urigüen and Garcia-Zapirain, 2015). It is relatively simple to remove artefacts from non-physiological sources. First, to keep the brain electrode in good contact with the scalp, a special EEG gel can be used. Asking participants to wash their hair in advance, or even cut their hair short will also help to keep the electrode in good contact. Furthermore, a ground wire can be connected with the earthing end of the instrument which was used to measure the EEG to restrain the power frequency interference. A reference electrode is also beneficial for restraining power frequency interference. However, it is not easy to handle the artefacts from physiological sources, because the noise sources are very diverse and have different characteristics. Therefore, physiological sources artefact removal is an important step in any EEG study. The current methods of physiological sources artefact removal are introduced below.

Literature on EEG artefacts removal is very extensive. However, so far researchers in the field have not agreed on the best ways to improve the quality of recorded signals. In early research, researchers always rejected the EEG epochs which were contaminated (Verleger, 1993; Kirkove, François, and Verly, 2014). However, if all contaminated epochs are rejected, much useful information will be lost too, with only a few usable

trials left (Croft and Barry, 2000b). The better choice is rejecting the epochs which contain excessive interference (Nolan, Whelan, and Reilly, 2010) and maintaining the rest using a specific method to remove the component of artefacts. Another option is to instruct participants to avoid eye blink and movement, and relax the muscles to avoid muscle movement, however, it will distort the EEG due to the relationship between cognition and eye movement (Verleger, 1991), and this is difficult to achieve over long periods of experimental recordings (Fatourechi *et al.*, 2007).

In recent years, a few novel methods have been proposed in the artefact removal area. The classic methods are: filtering (Sweeney, Ward, and McLoone, 2012), Morphological component analysis (MCA), and blind source separation (BSS) techniques, such as Principle Component Analysis (PCA) (Berg and Scherg, 1991), and Independent Component Analysis (ICA) (Makeig, Bell, Jung, and Sejnowski, 1996; Krishnaveni, Jayaraman, Kumar, Shivakumar, and Ramadoss, 2005; James and Hesse, 2005).

Filtering Filtering is a very practical way to remove several artefacts roughly. For example, the baseline bias is basically in the range of 0.1-1 Hz. Using a high pass filter can easily filter the baseline drift. Most of the Electrooculogram (EOG) activity is at frequencies below 4 Hz and EMG activity (such as movement of the head, body, jaw or tongue) at frequencies higher than 30 Hz (McFarland, McCane, David, and Wolpaw, 1997; Fatourechi *et al.*, 2007). Theoretically, a bandpass filter (4-30 Hz) can be applied to remove all EOG and EEG artefacts. However, there are also EEG components below 4 Hz and above 30 Hz, where EEG signals and artefacts overlap. In this situation, the simple bandpass, low pass and high pass filters are not suitable to remove the artefact, and other methods are required. A paper by Sweeney *et al.* (2012) presented many other filter methods:

- Adaptive Filtering: Adaptive filtering uses reference signals to estimate artefact signals and subtract the estimated signals to leave the estimation clean EEG signal (He, Wilson, and Russell, 2004). This method is easy to implement and can be used online, but it requires extra electrodes to record the reference signal to estimate the artefact signal, making the hardware more complex.

- Wiener Filtering: This does not require additional electrodes to record the reference signal. Therefore, in this respect, it is better than adaptive filtering. However, it cannot be used in online processing because of the need for prior calibration (Izzetoglu, Devaraj, Bunce, and Onaral, 2005).

- Bayes Filtering: A Bayes filter combines the advantages of both adaptive filtering and Wiener filtering. It does not require a reference electrode. Therefore, it reduces the hardware's complexity, and it is capable of operating in real-time processing. However, Bayes filter is computationally intractable. Therefore, Bayes filter needs other methods to implement the approximations of it, such as Kalman Filtering (Kalman, 1960), and Particle Filters (Candy, 2016).

Morphological component analysis (MCA) MCA is good for removing artefact but it requires pre-knowledge of the shapes of the artefacts (Matiko, Beeby, and Tudor, 2013; Yong, Ward, and Birch, 2009), which is often hard to know given individual differences. The shape of the eye blink signal can be directly recorded with additional electrodes placed close to the eye and then be subtracted from the EEG signal to reduce contamination (Croft and Barry, 2000a,c; Gratton, 1998; Gratton, Coles, and Donchin, 1983; Schlogl, Keinrath, Zimmermann, Scherer, Leeb, and Pfurtscheller, 2007). But, this blink-focussed recording increases set up time, and decreases channels available for EEG.

Blind source separation methods (BSS) Many advanced algorithms have been invented to remove artefacts in EEG. The blind source separation (BSS) method is one of them. This method has become increasingly common in the artefact removal field (e.g., Vigário, 1997; James and Hesse, 2005; Delorme, Palmer, Onton, Oostenveld, and Makeig, 2012; Daly, Nicolaou, Nasuto, and Warwick, 2013; Kim and Kim, 2018).

The basic concept of BSS is the separation of different source signals from raw signals without reference (Cardoso, 1998). Therefore, it does not need an extra electrode to record the reference signal to help to remove the artefacts because it separates the artefact components from all recorded EEG data.

There are different methods of blind signal separation:

- Principal components analysis (PCA): PCA was invented in 1901 by Pearson and was later independently developed and named by Hotelling in the 1930s. From the 1990s, PCA started to be used in EEG artefact removal (Berg and Scherg, 1991).

  PCA is a statistical procedure, and it uses orthogonal transformation to convert observations of possibly correlated variables into values of linearly uncorrelated variables. The variables are called principal components. In EEG artefact removal, raw EEG signals are separated into different components, and some

components are recognised as artefacts to be removed. However, PCA cannot completely remove artefacts from EEG, especially when they have comparable amplitudes (Jung, Makeig, Humphries, Lee, Mckeown, Iragui, and Sejnowski, 2000).

- Independent component analysis (ICA): ICA (Jung, Makeig, Westerfield, Townsend, Courchesne, and Sejnowski, 2000; Vigário, 1997) is a blind signal separation method to decompose signals into independent non-Gaussian signals. It was invented after PCA and was first introduced by Hérault and Ans in 1984. In 1994, Comon most clearly stated ICA. The first application in EEG was declared in an article by Makeig *et al.* (1996), and they proved that ICA could isolate a wide range of artefacts to remove. Other researchers also applied ICA in this area and approved that ICA is efficient in removing EEG artefacts (e.g., Jung *et al.*, 2000; Vigário, 1997). In practice, there are a few ICA algorithms such as SOBI (Belouchrani, Abed-Meraim, Cardoso, and Moulines, 1993), InfoMax (Lee, Girolami, and Sejnowski, 1999) and fastICA (Hyvärinen and Oja, 1997). One problem with the basic ICA function – as in the EEGLAB plug-in of MATLAB (Brunner, Delorme, and Makeig, 2013; Delorme and Makeig, 2004) – is that it only decomposes the raw data to a set of multiple independent components; an additional manual selection by an expert is needed to determine which component is related to artefacts. Not only is this choice process very time-consuming but sometimes the wrong component can be chosen for removal. The choice process has also been automated in both commercial software such as BrainVision Analyser (Brain Products, Germany) and in open source software such as ADJUST. ADJUST is an additional plug-in of EEGLAB, which classifies independent components in agreement with manual expert choice on 95% of the data variance (Mognon, Jovicich, Bruzzone, and Buiatti, 2011). Even coupled with ADJUST, ICA retains limitations: it not only removes EEG artefacts but it can also remove useful components as well (Congedo, Gouy-Pailler, and Jutten, 2008; Inuso, La Foresta, Mammone, and Morabito, 2007); and ICA is normally used to separate N components from N channels (Jung *et al.*, 2000), can not simply work with a single channel.

- Others: There are other BSS methods such as non-negative matrix factorization (Lee and Seung, 2001), low-complexity coding and decoding (Hochreiter and Schmidhuber, 1997) and stationary subspace analysis (Zeng, Song, Yan, and

Qin, 2013) which can also be used for EEG artefact removals.

The BSS methods are now widely used in EEG artefacts removal because they do not require extra electrodes to record artefact reference signals. However, BSS methods have their own limitations. The compositions of signals to be decomposed should be uncorrelated, independent, have non-Gaussianity, instantaneous propagation, and linearity, etc. Otherwise, the effect will not be good. Meanwhile, in EEG artefacts removal, if there is more than one artefact (N artefacts) to separate, at least $N + 1$ channels are needed to record the signal. The biggest common disadvantage of BSS methods is that they cannot separate the artefacts' components precisely, so it either leads to a residual of the artefacts' components or the EEG signal itself being removed.

Conclusion The EEG artefacts removal methods were introduced above. These methods all have their own advantages and disadvantages. The ICA algorithm is the most popular method in recent years as it has been reported to be successful for all kinds of contaminants (Urigüen and Garcia-Zapirain, 2015).

In this thesis, several methods were combined to clean our experimental data. Furthermore, a novel eye blink removal method based on physiology was invented (Zhang *et al.*, 2017). ICA was adopted as a golden standard to compare with the method we proposed. The results of our method are better in three aspects: it can be easily used with both single- and multi-channel EEG; it does not change the pure EEG data; it should allow online processing for experiments such as neurofeedback training.

2.4.3.2 EEG feature extraction methods

The interconnection of vast numbers of neurons makes EEG signals highly complex, unstable and chaotic. Extracting and collecting features invisible to the naked eye from raw EEG is an important part of EEG analysis. The key step of EEG feature extraction is called signal processing. The popular signal processing algorithms used in EEG analysis are introduced below.

Time-frequency domain analysis Time-frequency domain methods are the most commonly used EEG signal analysis methods. The time domain signal is rarely used to detect EEG features directly (Bhat *et al.*, 2014) because of the irregular nature of EEG. Therefore, the signals are always transferred into the frequency domain, but in some cases, they do not keep the time information. With the progress of technology, the time and frequency information can both be kept to reflect the features of EEG.

- Frequency domain: Frequency domain analysis of the EEG has had a long history. The first frequency analysis in EEG was in 1932 when Berger mentioned that some of the EEG he had recorded had undergone frequency analysis by Dietsch. After that, more researchers use EEG's frequency domain information to analyse EEG data. In particular, after the fast Fourier transform algorithm (FFT)'s (Cooley and Tukey, 1965) invention and the increasing of computing power, the frequency domain analysis in EEG is widely used.

  FFT is a very useful tool that can present the dominant frequency components in a signal (Subhash Chandran *et al.*, 2016), and the obtained EEG power through FFT is a very important feature in physiological status detection.

  In 1973, Gotman *et al.* calculated the EEG powers in different frequency bands (delta, theta, alpha, and beta) to study some of the relations between these activities. Brenner, Ulrich, Spiker, Sclabassi, Reynolds, Marin, and Boller (1986) analysed EEG power of major depression patients and healthy elderly controls. Compared to the controls, the parasagittal mean frequency and beta activity of the depressed patients were decreased. Experiments designed by Nofzinger, Price, Meltzer, Buysse, Villemagne, Miewald, Sembrat, Steppe, and Kupfer (2000) clarified the neurobiological basis of variations in one aspect of central nervous system 'arousal' in depression by characterising the functional neuroanatomic correlates of beta EEG power density during non-rapid eye movement (NREM) sleep. The results showed that in the depressed group, there was a trend for beta power to correlate with an indirect measure of absolute whole brain metabolism during NREM sleep. Grin-Yatsenko, Baas, Ponomarev, and Kropotov (2009) compared EEG spectrum power obtained in the early stages of depressed patients with age-matched healthy controls. There was a significant increase in spectrum power in the theta, alpha , and beta frequency bands of depressed patients at the parietal and occipital sites. Knott, Mahoney, Kennedy, and Evans (2001) also suggested the EEG power as a useful tool for investigating brain regional mechanism in depressed patients. The traditional power, inter-hemispheric power ratio, intra-hemispheric power ratios, mean frequency and both inter- and intra-hemispheric coherence indices were all computed to be compared between depressed male patients and healthy controls. The results showed that depressed patients have greater overall relative beta power, greater absolute beta power and faster mean total spectrum frequency in the bilateral anterior regions and intra-hemispheric theta power asymmetry reduction in all regions. The classification accuracy of

classifying the patients and controls through the EEG power achieved 91.3%.

- Time-Frequency domain: Subhash Chandran *et al.* (2016) mentioned that brain signals change rapidly. For example, once the brain received a visual stimulus, all visual areas became activated in less than 150 ms (Schmolesky, Wang, Hanes, Thompson, Leutgeb, Schall, and Leventhal, 1998; Tamura and Tanaka, 2001). Taking this fact into consideration, spectral analysis should be done in short time intervals of a signal called time-frequency analysis. There are several techniques to realise time-frequency analysis, such as short time Fourier transform (STFT) and wavelet transform.

    1. STFT: EEG signal is unstable because it changes in a short time frame, whereas STFT can be applied to process the signal and get the EEG power in a short time. It applies window functions in the signal, then transfers the short signal into frequency domain. The time-frequency resolution is fixed (Suman, Sumabindu, and Seventline, 2015).

        Sheikhani, Behnam, Mohammadi, and Noroozian (2007) investigated the EEG background activity in patients with Autism, and they calculated short time Fourier transform (STFT) and STFT at the bandwidth of the total spectrum (STFT-BW) of the EEG in 10 Autism patients and 7 controls. They found a significant difference of STFT-BW between these two groups. Using this feature to discriminate the two groups, the accuracy was 82.4%. Behnam, Sheikhani, Mohammadi, Noroozian, and Golabi (2008) assessed the STFT value of EEG with variance analysis in 10 Autism disorders and 9 age-matched control subjects. Results showed that the alpha band had 89.5% discrimination between the two groups. Tzallas, Tsipouras, and Fotiadis (2009) calculated the STFT of EEG in seizure patients and control groups. The value of STFT then fed into an artificial neural network, and the accuracy of discrimination of normal and seizure is up to 99.8%.

        However, STFT has a disadvantage in which the length of the window is finite. Wide windows provide good frequency resolution and relatively poor time resolution, whereas narrow windows provide good time resolution and relatively poor frequency resolution (Subha *et al.*, 2010). We have to find a balance to set the length of window. The length's setting is empirical and not objective enough. An inappropriate length of window will influence the result of time-frequency analysis. However, this resolution problem can be

solved by wavelet transform (Subha *et al.*, 2010).

2. Wavelet: Wavelet transform theory has many applications in digital signal processing (Mallat, 1989).

   Hazarika, Chen, Tsoi, and Sergejew (1997) described an application of wavelet transform. It uses a wavelet transform-based feature in EEG signals to classify three kinds of EEG signals: normal, schizophrenia (SCH), and obsessive compulsive disorder (OCD). The network with wavelet coefficients was able to correctly classify over 66% of the normal class and 71% of the schizophrenia class of EEGs, respectively. Subasi (2007) investigated the diagnosis of epilepsy, using wavelet feature extraction and a mixture of expert model based on the EEG signal. Raw EEG signals were decomposed into the frequency sub-bands using discrete wavelet transform (DWT), and the sub-band frequencies were used as an input to an ME network with two discrete outputs: normal and epileptic. The highest correct classification accuracy of the EEG signals was up to 95%. In a paper (Ahmadlou, Adeli, and Adeli, 2010, 2012b), two wavelet-based methodologies were presented for a diagnosis of autistic spectrum disorder. They both had high classification accuracy of up to 90% and 95.5% for distinguishing autistic EEGs from healthy EEGs. To predict antidepressant's treatment outcomes based on EEG, a machine learning method based on a feature matrix which was constructed involving time-frequency decomposition of EEG data based on wavelet transform analysis was proposed. The wavelet transform analysis has shown high classification accuracy, i.e., accuracy = 87.5%, sensitivity = 95%, and specificity = 80% (Mumtaz, Xia, Yasin, Ali, and Malik, 2017). Puthankattil and Joseph (2012a) attempted to classify the EEG signals of normal and depression patients using relative wavelet energy and artificial feedforward neural network. The performance of the network was evaluated using the classification accuracy and its value of 98.11%, and it indicated that it has great potential for classifying normal and depression signals.

Other methods Except for time-frequency domain methods, researchers adopted various other signal processing algorithms to extract features from complex and unstable EEG signals :

- Fractal dimension (FD): The term 'fractal dimension' was created by Mandelbrot in 1967. After that, fractal theory was used to examine biomedical signals, and it

can measure their complexity (Mandelbrot, 1982; Barnsley, 2014). It is a powerful tool to detect the transients in signals (Raghavendra and Dutt, 2010). Therefore, some researchers adopted this method to analyse EEG signals (e.g., Pradhan and Dutt, 1993; Woyshville and Calabrese, 1994; Ahmadlou *et al.*, 2010; Ahmadlou, Adeli, and Adeli, 2012a; Wang, Zhou, Yuan, Li, Meng, Zhao, and Wang, 2013; Hsu, 2013; Wang *et al.*, 2013). However, FD has limitations. It is highly sensitive to the sampling frequency, and it is not suitable for time series waveforms because the interval along the X axis (time) cannot compare with that along the Y-axis (amplitude) (Raghavendra and Dutt, 2010).

- Recurrence quantification analysis (RQA): RQA was developed to quantify differently appearing recurrence plots (RPs), whereas RPs are tools which can be used to visualise the recurrence of signal states in phase space (Marwan, Thiel, and Nowaczyk, 2002; Marwan, Romano, Thiel, and Kurths, 2007). This method is good to analyse the non-stable signal because it can find the number and duration of recurrences in a dynamical system (Acharya, Sree, Chattopadhyay, Yu, and Ang, 2011). Therefore, it is used in some EEG research for various purposes (e.g., Pijn, Velis, van der Heyden, DeGoede, van Veelen, and da Silva, 1997; Song, Lee, and Kim, 2004; Acharya *et al.*, 2011).

- Higher-order spectra (HOS): The higher-order spectrum is an extension of the Fourier spectrum, which represents the moments and cumulants of the signal (Chua, Chandran, Acharya, and Lim, 2008b). Therefore, it can find the nonlinear characteristics in the complex patterns of ECG (e.g., Chua *et al.*, 2008b; Chua, Chandran, Acharya, and Min, 2009), EEG (e.g., Chua, Chandran, Acharya, and Lim, 2008a, 2009; Pradhan, Jena, Nadar, and Pradhan, 2012; Muthuswamy, Sherman, and Thakor, 1999). The representations of HOS are bispectrum and bicoherence (Subha *et al.*, 2010).

- Entropy: Entropy is the rate of information production (Richman and Moorman, 2000). However, estimations of entropy are not well suited to short and noisy data (Richman and Moorman, 2000) (such as EEG). Therefore, Pincus (1991) devised approximate entropy (ApEn) which was suitable to analyse biological signals (Pincus, 1991, 1995). This method has been widely applied in EEG studies (Pincus and Keefe, 1992; Abásolo, Hornero, Espino, Poza, Sánchez, and de la Rosa, 2005; Srinivasan, Eswaran, and Sriraam, 2007; Ocak, 2009; Puthankattil and Joseph, 2014; Faust *et al.*, 2014). However, ApEn also has its limita-

tions because it relies on the record lengths and it lacks relative consistency (Pincus, 1995). Another entropy-based method called sample entropy(Samp-En) was developed which overcomes the shortcoming of ApEn, and does not rely on the length of record and shows consistency which ApEn does not have. It also needs less calculating time than ApEn (Richman and Moorman, 2000). Therefore, Samp-En is used in many EEG studies (Song, Crowcroft, and Zhang, 2012; Zhang, Wang, and Wu, 2013; Jie, Cao, and Li, 2014).

- Lyapunov exponent (LE): Lyapunov exponents estimate the mean exponential expansion or contraction of flow in phase space (Wolf, Swift, Swinney, and Vastano, 1985). Fell, Röschke, and Beckmann (1993) used LE to distinguish the EEG of different sleep stages, and the result shows the significant difference in different stages. Kim, Jeong, Chae, Park, Kim, Go, Paik, Kim, and Choi (2000) used LE to compare schizophrenic patients' and controls' EEG. The results showed that schizophrenic patients had lower LE values in the left inferior frontal and anterior temporal regions compared with normal controls. Krystal, Zaidman, Greenside, Weiner, and Coffey (1997) calculated the LE of EEG to classify seizures and reflect the relative physiological impact of seizure activity.

- Conclusion: There are many other signal processing algorithms used in EEG analysis, such as Hurst's exponent (HE) (Dangel, Meier, Moser, Plibersek, and Shen, 1999), Detrended Fluctuation Analysis (DFA) (Lee, Kim, Kim, Park, and Kim, 2002), wavelet-chaos methodology (Adeli, Ghoshdastidar, and Dadmehr, 2010), Wavelet Packet Decomposition (WPD) entropy (Jiang and Adeli, 2004; Jiang, Mahadevan, and Adeli, 2007; Su, Huang, Chen, Liu, Huang, and Le, 2014) and so on. They all have their advantages and disadvantages.

In our research, we used two time-frequency analysis methods: FFT and wavelet to calculate the important information — EEG power. Another spatial filter method — CSP was also adopted. This method is usually used in motor-imaginary EEG classification, seldom used in emotional EEG detection. Because of its good spatial distinguishability, we used it in our research. The character of CSP is introduced in Section 5.3.5 of Chapter 5.

2.4.3.3 EEG feature analysis

The traditional method to analyse the features extracted from EEG is statistical analysis, and the most commonly used is analysis of variance (ANOVA). For example,

this method is used to classify depressed patients' and normal controls' EEG features (Debener *et al.*, 2000; Knott *et al.*, 2001), Alzheimer's and normal controls' EEG feature (Abásolo *et al.*, 2005), and test the EEG coherence in Alzheimer's disease (Besthorn, Förstl, Geiger-Kabisch, Sattel, Gasser, and Schreiter-Gasser, 1994).

In recent years, machine learning methods have become more popular in EEG analysis, such as Support Vector Machine (SVM) (Vapnik, 1998; Burges, 1998; Acharya, Sudarshan, Adeli, Santhosh, Koh, Puthankatti, and Adeli, 2015; Frid and Manevitz, 2018; Frid, 2014), Decision Tree (DT) (Han and Kamber, 2005), K-Nearest Neighbour (Han and Kamber, 2005), Artificial Neural Networks (ANN) (Adeli and Hung, 1995; Adeli and Park, 1998; Siddique and Adeli, 2013; Pritchard, Duke, Coburn, Moore, Tucker, Jann, and Hostetler, 1994), enhanced probabilistic neural networks (Adeli and Panakkat, 2009; Ahmadlou and Adeli, 2010; Sankari and Adeli, 2011), Probabilistic Neural Network (PNN) classifier (Faust *et al.*, 2014) and so on.

The choice of feature analysis methods depends on the purpose of the experiments and the feature extracted from EEG. Different feature analysis methods also supply different results. Table 2.2 summarises some combinations of different feature extraction methods and feature analysis methods in different application fields. Because almost all related papers used EEG to classify mental-disorder patients and controls, the following table lists the cases where classification results are over 90%.

Application	**Paper**	**Feature extraction**	**Feature Analysis**	**Results**
autism	(Ahmadlou *et al.*, 2010)	Wavelet decomposition, HFD, KFD	RBFNN	90%
autism	(Behnam *et al.*, 2008)	STFT-BW + coherence measures	KNN	94.40%
autism	(Ahmadlou *et al.*, 2012b)	Wavelet decomposition, fuzzy SL	EPNN	95.50%
Alzheimer's	(Pritchard *et al.*, 1994)	D2	ANN	92.3%
depression	(Bachmann *et al.*, 2017)	SASI+DFA	LDA	91.2%
depression	(Knott *et al.*, 2001)	power, frequency, asymmetry, coherence measures	Statistical analysis	91.3%
depression	(Ahmadlou *et al.*, 2012a)	Wavelet decomposition, HFD, KFD	EPNN	91.3%

depression	(Acharya *et al.*, 2015)	FD, LLE, SampEn, DFA, Hurst's exponent, HOS, RQA	SVM, KNN, NBC, PNN, DT	98%
depression	(Puthankattil and Joseph, 2012b)	RWE	ANN	98.11%
depression	(Faust *et al.*, 2014)	WPD, ApEn, SampEn, REN, Ph	PNN, SVM, DT, kNN, NBC, GMM, FSC	99.5%
epileptic	(Acharya *et al.*, 2011)	RQA	SVM, GMM, FSC, KNN, NBC, DT,	95.6%
epileptic	(Wang *et al.*, 2013)	FD	SVM	97.58%
epilepsy	(Chua *et al.*, 2009)	HOS	GMM, SVM	93.11%
epilepsy	(Srinivasan *et al.*, 2007)	ApEn	ENN,PNN	100%

Table 2.2: Summary of research on EEG studies

From Table 2.2, we can see that the machine learning methods applied to EEG analysis have achieved very high accuracy in classification. However, it is not enough to diagnose depression as a binary disease, as the severity of depression (depressivity) is very important. It is also not enough to develop black-box biomarkers that are not easily interpretable biologically, as these will not be well accepted amongst clinicians.

In this research, we used AdaBoost, LSBoost (Friedman, 2001), random forest, lasso and one of the neural networks — LVQ to analyse the features extracted from the EEG signal to predict depressivity and also to highlight potential interpretable biomarkers extracted from EEG signals. These techniques are introduced in Chapter 6. Importantly, our work differs from previous work in that we predicted the continuous depressivity of each participant. This work is new, except for the recent report by Stewart and Allen (2018), with which we have compared our results. The bulk of previous work reported classification accuracies to evaluate classification performance, while we used percent variance accounted for (R^2-value) to evaluate regression performance.

2.4.4 Analytical Considerations

Apart from the issues mentioned above, variant variables such as participants' choice and sensor location are also important information for EEG analysis. Therefore, in this section, all these variables are discussed.

Gender Females have a higher incidence of depression than males (Acharya *et al.*, 2015). Due to the different incidences and causes of depression for different genders, researchers tried to find if there is a difference between females' and males' EEG behaviour linked to depression. Bryden (1982) reported that gender may modulate hemispheric EEG asymmetry, while Smit, Posthuma, Boomsma, and De Geus (2007) presented that EEG asymmetry may only relate to depression in young females rather than males. Also, research found that females' hemispheric asymmetry responded to Fluoxetine treatment, but males did not (Bruder, Stewart, Tenke, McGrath, Leite, Bhattacharya, and Quitkin, 2001). Under the usage of lithium, females and males also showed clear differences in the changing of EEG power (Thau, Rappelsberger, Lovrek, Petsche, Simhandl, and Topitz, 1988). The above research shows that EEG behaviour differs between genders. Thus, the male and female groups deserve to be studied separately.

Age Neurophysiological states vary with ageing, and the EEG signal is also affected by age (Acharya *et al.*, 2013). Therefore, the participants' age range in the experiments needs to match. In our research, we chose participants aged between 18-37 years, with a mean of 21.56 years.

Medication Studies have shown that medications will alter the characteristics of EEG, and the effects on the EEG are different and dose-dependent (Banoczi, 2005). The increase of theta and excess beta activity may be the most regular EEG alterations associated with medication (Blume, 2006). Thau *et al.* (1988) showed that after ten days of lithium intake, the absolute power of EEG was increased, especially in the theta and beta band. Meanwhile, the power increased in the left hemisphere and decreased in the right. Antidepressants with thymoleptic properties will cause the decline of alpha activity (Saletu, Grünberger, Rajna, and Karobath, 1980), while antidepressants without thymoleptic cause more distinct EEG pattern (Acharya *et al.*, 2015). Drugs also cause other changes such as diffuse delta, triphasic waves, bisynchronous spikes or polyspikes (Blume, 2006). Therefore, our research used drug-free participants.

Epoch length The collected EEG signals used for analysis is always a long period recording. Based on analysis requirements, a long period of recording will be split into short epochs. Previous research showed that epoch-to-epoch variability of power spectra computed with long epoch lengths, e.g. 8, 16, and 32-seconds, was significantly higher than those computed with short epochs, e.g. 1, 2-seconds (Levy, 1987; Allen, Coan, and Nazarian, 2004). We used the fast Fourier transform (FFT) as one of the signal processing methods. The FFT requires 2^n-data-point length epoch to be analysed (Allen *et al.*, 2004). Since our sampling rate was a power of two (256 Hz), we used 1-second epochs in most of our models (Epoch splitting is introduced in Section 5.2).

Handedness Previous research found significant differences in EEG activity between the right-handed and left-handed groups of subjects (Provins and Cunliffe, 1972). Some studies on personality correlations with EEG measures even restrict their sample to right-handed women to reduce sources of conflicting variance (Shackman, McMenamin, Maxwell, Greischar, and Davidson, 2009). Therefore, in our research, we only used right-handed participants' data.

2.5 Conclusion

This chapter introduced the psychological terms 'depression' and 'depressivity'; and discussed their definition, diagnosis and the difficulties encountered in diagnosis. The human brain's structure was also introduced to supply a physiological basis for this study. EEG as the most important research object was also introduced. The related work of EEG analysis, especially on different physiological states was also discussed. The literature review has provided a guide to advance our research.

Chapter 3

The Details of the EEG Dataset and an Overview of the Depressivity Prediction Model

In this thesis, the primary goal is to use machine learning methods to predict the depressivity of participants. In this chapter, an overview of the depressivity prediction models' design is introduced. The EEG dataset which was used in this project is also described in this chapter. This dataset includes participants' basic personality and demographic information and their EEG data. In addition, the data preprocessing steps are also introduced.

3.1 Introduction

The dataset reported in this chapter was used in the whole research (apart from Chapter 4), and it is the basis of the thesis. This dataset includes the participants' details such as gender, age and their resting EEG data. It also includes the depressivity score (PID5-d) obtained by a questionnaire. Because the dataset had some errors and omissions, before the main experiments, we sorted the dataset in advance. This is an important step in the research because the data are the key information in the study.

The raw data were inevitably contaminated by a lot of artefacts. Therefore, they needed to be preprocessed before they were analysed. These preprocessing steps are also introduced in this chapter.

The basic steps of the depressivity prediction models are introduced in this chapter to provide an overview of the models.

3.2 EEG Dataset

Data for this project were obtained from an existing dataset, the characteristics of which are described in the following sections. This dataset was used for this whole project, except for Chapter 4.

3.2.1 Participants

The original dataset included 83 participants, but 10 of them did not have adequate data (See Figure 3.2, the verified process will be discussed later). Therefore, there were 73 participants' data samples (44 females, 29 males; aged 18-37 years with a mean of 21.56 years) left. All procedures were approved by the University of Otago Ethics Committee (approval number: H15/005). All participants were recruited through the University of Otago Student Job Search. Before participating in the experiment, the participants were provided informed consent. Each participant received NZ$15 per hour for their time and effort spent in the experiment. Participants reported no medical or psychological treatment for depression, anxiety or other types of emotional disorder in the last 12 months. All participants were right-handed.

3.2.2 Procedure

Prior to the EEG testing, participants completed the first part of a computer-delivered questionnaire program. This questionnaire program contains the Spielberger State-Trait Anxiety Inventory Y-form (Spielberger and Gorsuch, 1983) Trait Anxiety scale items, the Eysenck Personality Questionnaire-Revised Extraversion and Neuroticism scale items (Eysenck, 1991) and the Behavioural Activation System/Behavioural Inhibition System (Carver and White, 1994) BIS scale items (duration 10-15 minutes). The experimenter then measured each participant's head circumference and marked Fp1 and Fp2 according to the International 10-20 system using a black marker. Participants were fitted with the appropriately sized Wave guard (Ag/Agcl) EEG cap which was then connected to a 32-channel ANT system (Advanced Neuro Technology B.V., Enschede, The Netherlands). The sampling rate for analysis was 256 Hz. Electrode gel (Electro-Cap International, Eaton, OH, USA) was inserted into each electrode via a blunt square-tipped 16-gauge needle (Precision Glide, Needle, Becton Dickinson, Franklin Lakes, NJ, USA); impedance was reduced to <5KΩ by gentle abrasion of the scalp with the tip of the needle; and a brief relaxation-induced alpha rhythm and deliberate eye blink traces assessed by the experimenter to ensure a good recording, with

adjustments as necessary (preparation time ~ 30 minutes). A relaxation test was then performed by instructing the participants to remain relaxed, with their eyes open and then closed for one minute intervals, the order was O-C-O-C-C-O-C-O (C: eyes close, O: eyes open). Their EEG were recorded throughout this period (8 minutes' duration).

In the procedure, 19/73 participants were recorded at 32 channels: 'Fp1', 'Fpz', 'Fp2', 'F7', 'F3', 'Fz', 'F4', 'F8', 'FC5', 'FC1', 'FC2', 'FC6', 'A1', 'T7', 'C3', 'Cz', 'C4', 'T8', 'A2', 'CP5', 'CP1', 'CP2', 'CP6', 'P7', 'P3', 'Pz', 'P4', 'P8', 'POz', 'O1', 'Oz', 'O2'. However, others were recorded only at 20 channels: 'Fp1', 'Fpz', 'Fp2', 'F7', 'F3', 'Fz', 'F4', 'F8', 'A1', 'T7', 'C3', 'Cz', 'C4', 'T8', 'A2', 'P7', 'P3', 'Pz', 'P4' and 'P8'. To keep each data sample comparable, redundant channels were removed from the 32-channel data samples, leaving 20 channel's data. 'A1' and 'A2' were recorded as reference channels, so these two channels were also excluded. After that, each data sample included 18 channels. Figure 2.2 shows the positions of these channels. The channels which were used are marked by red circles, and reference channels are marked by blue circles. The data were referenced to CPz when recording, and the average of 'A1' and 'A2' was used to re-reference.

3.2.3 Pid-5 Depressivity Scores (PID5-d Scores)

After EEG testing, all analysed participants completed a set of scales from the questionnaire of 'Personality Inventory for DSM-5'. The depressivity scores (PID5-d) for each participant were calculated. Because the depressivity scale includes 14 items with a 4-point, which we scored as 0-3, respectively, the highest possible score is 42, and the lowest score is 0. In our dataset, the highest PID5-d is 27, and the lowest is 0. The score represents the depressivity of each participant. Figure 3.1 shows the distribution of female group and male group scores. There are more low PID5-d participants than high score participants because the participants of this dataset are not clinical patients.

3.2.4 Dataset Tidying

As Section 3.2.1 mentions 10/83 participants' original data were not adequate. The detection and tidying steps are presented in this part.

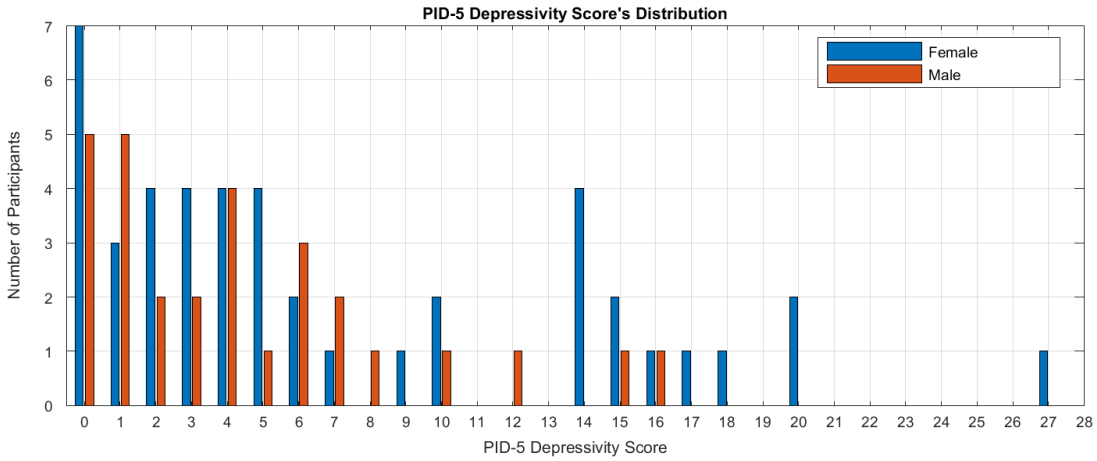

Figure 3.1: The figure illustrates the PID-5 depressivity scores' distribution of our data samples. We can see that most data samples are located in the low score section, and a small number of data samples are located in the high score section.

3.2.4.1 Check PID5-d scores

Participants were asked to complete a computer-delivered questionnaire program to obtain their PID5-d scores before EEG testing. Due to various reasons, six participants failed to complete the questionnaire form. Therefore, these six data samples had to be removed from the dataset. There were 77 data samples left (Figure 3.2).

3.2.4.2 Check the length of resting records

Each participant should have 8 minutes' resting EEG data. The sample rate was 256 Hz. Therefore, 122880-points-length resting segments would match our expectations ($8 \times 60 \times 256 = 122880$). However, four data samples in the dataset did not meet the requirement. The resting segments' lengths of these four data samples were not around 122880. The possible reason is that they had wrong labels to indicate the beginning and end of the resting records. Therefore, these four data samples were also removed from the dataset. After that, there were 73 data samples left.

3.2.4.3 Check the records of the reference channels

In all data samples, 'A1' and 'A2' were reference channels. Their average is used to reference all the data, so their values should be equal and their signs are opposite. All the data samples were checked, and the results show that all 73 data samples meet the rules. Therefore, after that step, no data sample was excluded (Figure 3.2).

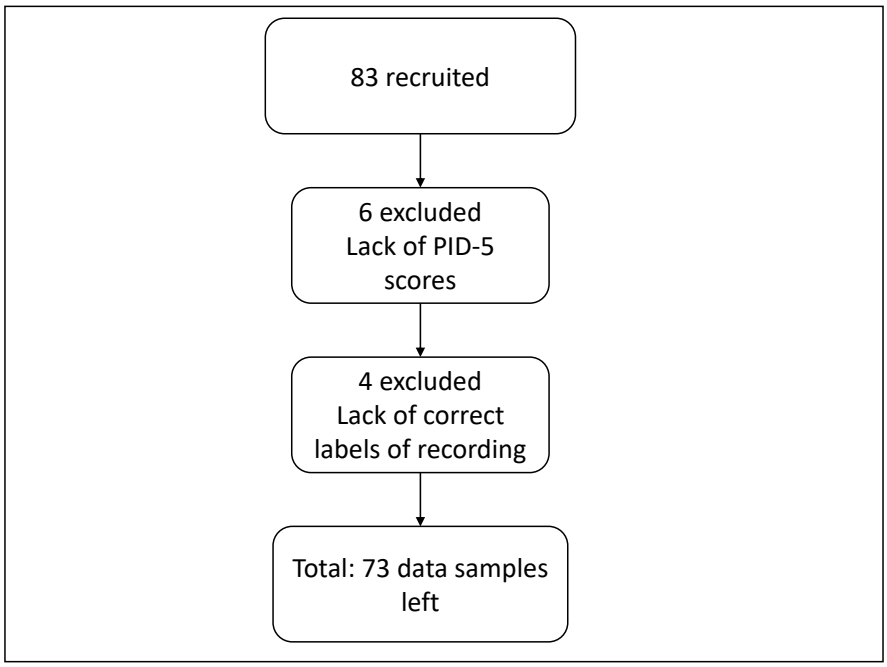

Figure 3.2: The figure illustrates that after dataset tidying, 10 data samples were excluded.

3.3 Overview of Depressivity Prediction Model

The dataset was then used to establish the depressivity prediction models. The dataset does not include a diagnostic class of depression patients; we have the depressivity score of each participant in our project. Therefore, our aim is not to distinguish a 0/1 classification (depression or not depression), but to get a numerical estimate of the depressivity of participants. Therefore, several depressivity prediction models using machine learning methods were designed to generate a machine learning depressivity (ML-d) score as a prediction of the measured trait depressivity (PID5-d) of the participants. Then, the correlation coefficient (*R-value*) between the ML-d score and the PID5-d score was calculated to evaluate the models (with $R^2 \times 100\%$ measuring the percentage of variance accounted for).

Although we needed to predict a numerical value, the first kind of model we tried was classification-based. Its design idea is based on the assumption that the frequency of occurrence of some specific EEG patterns that indicate High depressivity increases with a higher depressivity. It means that each epoch of a signal is represented by an EEG pattern to indicate this epoch belongs to High depressivity or Low depressivity.

If a participant has more High depressivity pattern epochs, it means the participant is highly depressed. If a participant has less High depressivity epochs, it means the participant is less depressed, and more healthy. Therefore, in classification-based models, the main aim is to distinguish the epochs as a high (High depressivity) or low (Low depressivity) class. Depressivity is estimated by the number of high epochs in all epochs (Equation 3.2). The second kind of models we tried were regression-based models. The design idea is based on an assumption that EEG patterns vary with different depressivity.

According to different design ideas, these two kinds of models adopted different predictive analysis methods.

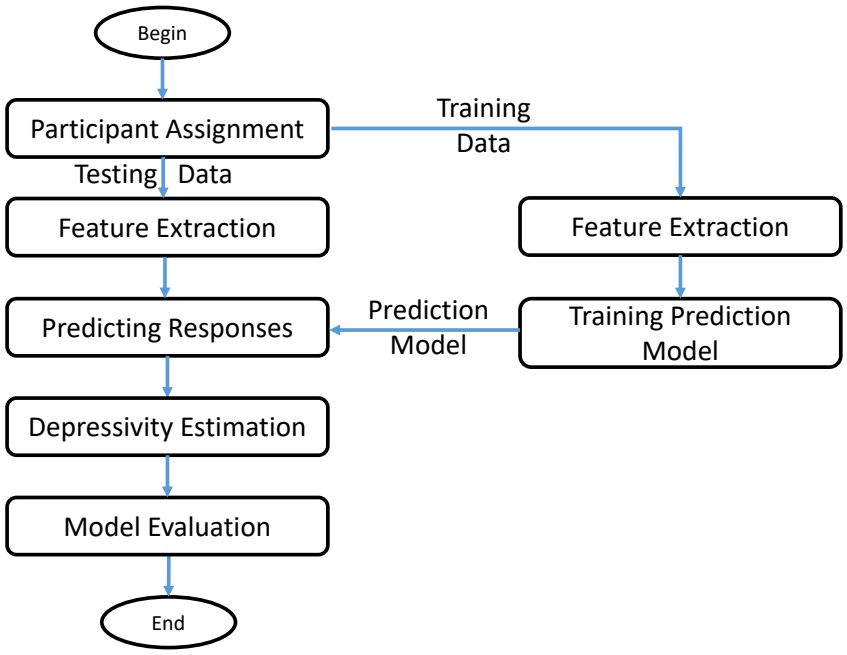

Figure 3.3: Steps of the establishment and evaluation of depressivity prediction models: participant assignment, feature extraction, predictive analysis, depressivity estimation and model evaluation. At the participant assignment step, data were separated into two parts: training data and testing data. Training data were used to train prediction models. Features extracted from each participant in the testing data set were fed into the prediction model to obtain a predicted value for each participant.

Figure 3.3 illustrates the basic steps of the establishment and evaluation of depressivity prediction models. Both classification- and regression-based models follow these steps. They are the participant assignment, feature extraction, predictive analysis, depressivity estimation and model evaluation.

3.3.1 participant assignment

As Figure 3.3 shows, in depressivity prediction models, all data samples were separated into two parts: training data and testing data. Training data were pooled together to obtain feature parameters and establish a prediction model. From the testing data group, features were extracted and analysed by the prediction model for each participant. Different assignment methods were used in classification-based models and regression-based models separately. They are introduced in the following sections.

3.3.1.1 Classification-based models

This section introduces the participant assignment in the classification-based model.

As mentioned in section 3.2, we took an existing dataset of 73 participants' EEG data and calculated their PID5-d scores. Figure 3.1 shows the distribution of PID5-d scores for our data samples. It shows that the distribution of the scores is imbalanced. For example, there are 12 data samples (7 females and 5 males) with a PID5-d score of 0, but there is only 1 data sample (female) with a score of 27.

In conventional methods, for data samples in which the classes are balanced, each score is set as one class and then x% data samples of each class are randomly chosen as training data and the rest as testing data. Usually, x is set as 50, 75, or 90, it depends on the number of data samples and the designing of the model.

In our project, there are many different scores in the dataset and they are imbalanced, so we required a new assignment method. Although there is a continuous range of different scores in the data sample, for training purposes, we only set two training classes: high (High depressivity) and low (Low depressivity).

We suppose that the same score data samples have the same performance. To avoid the very highest scores and very lowest scores being lost from testing, a training data selection method was designed to pick these two classes' participants. The method uses an extended 'ABBA' counterbalanced sequence: ABBABAABBAAB. The 'ABBA' counterbalanced sequence is usually used to control experiments' order (Shuttleworth, 2009) to eliminate sequence effects (Salkind, 2010). In this research, the extended counterbalanced sequence was used to select training data samples.

Figure 3.4 illustrates the ABBABAABBAAB sequence and the application in this research. In each experiment, all data samples were ranked by the PID5-d score. The highest one was at the front, and the lowest one was at the end. In round A, we selected data samples with high scores from the front with the A samples (6 samples)

and data samples with low scores from the end with the A samples (6 samples) for training. The B data samples were retained for testing. The intermediate participants left were also pooled into the testing group. Vice versa, in round B, B data samples were selected as training samples, and A data samples and the rest were used as testing sets. After round A and round B, A data samples and B data samples were tested only once, and intermediate samples were tested twice. Therefore, in the step of depressivity estimation (Section 3.3.3), the ML-d scores of the intermediate samples were the average of two rounds.

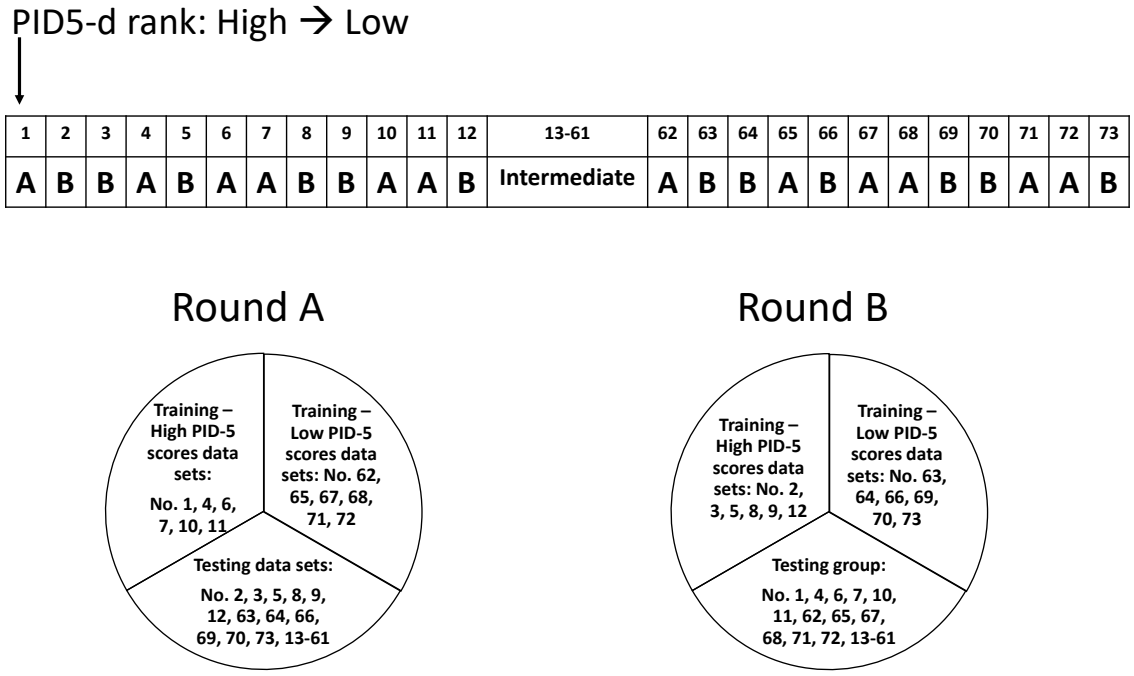

Figure 3.4: This figure illustrates the ABBABAABBAAB sequence and the application in this research.

The reason we used the ABBABAABBAAB sequence to select training sets is that if the top 6 samples and the bottom 6 samples were chosen as training data, the very highest scores and the very lowest scores would be lost. With the ABBABAABBAAB sequence, this problem was solved because a wider range of PID5-d scores' samples were left. We suppose that the same score case has the same performance, so the ABBABAABBAAB sequence method is good to balance the training and testing data group.

Because of the limitation of the number of data samples, in classification models, male and female were not split, but different eye states were split into different cate-

gories. Because if we split them into male groups and female groups, it will lead to a big difference of PID5-d scores in each group and a small gap between high and low group. For example, for male group, there are only 29 participants (Figure 3.1). We selected top A cases as high training group (PID5-d: 4, 6, 6, 7, 12,15) and bottom A cases as low training group (PID5-d: 0, 0, 1, 1, 1,2). The difference between the highest PID5-d score and the lowest PID5-d score in the high training group is 11 (15-4), while the difference between the highest score in low training group and the lowest score in high training group is only 2 (4-2). This split is inappropriate. Therefore, in classification-based models, male and female were not split.

3.3.1.2 Regression-based models

In regression-based models, to derive a more accurate estimation of prediction performance, k-fold cross validation was used to pick training and testing data samples. All participants (data samples) were split into k subsets, and each subset contained M participants.

$$M = \frac{T}{k} \tag{3.1}$$

where T refers to the number of all participants in the experiment. In each round, one subset was left to test, and other $k-1$ subsets were used to train the model. The cross validation was repeated for k rounds, with each participant tested once.

Figure 3.5 shows an example: 7-fold cross validation of male group. The male group had 29 participants, and these participants were split into 7 subsets. Each subset contained 4 participants. Note because there were 29 participants, the last subset contained 5 participants. Participants were labelled as 'P01', 'P02', etc. The number only indicates the participants' collection order. There is no relationship between the number and the PID5-d score or other data characters. In the first round, participants 'P05'-'P29' were used to train the sub-model. Participants 'P01'-'P04' were used to test the sub-model and obtained ML-d scores. In the second round, 'P05'-'P08' were used as testing participants, while others were used as training participants. After 7 rounds, each participant was tested once. Each epoch of each participant obtained a prediction value.

Because our dataset includes a lot of information, the participants also need to be split into different categories:

- Gender: For depressive disorder, gender difference exists (Piccinelli and Wilkin-

MAN: 7-fold

All Data	P01-04	P05-08	P09-12	P13-16	P17-20	P21-24	P25-29
Round 1	Testing	Training					
Round 2	Training	Testing	Training				
Round 3	Training		Testing	Training			
Round 4	Training			Testing	Training		
Round 5	Training				Testing	Training	
Round 6	Training					Testing	Training
Round 7	Training						Testing

Figure 3.5: This figure shows the example of 7-fold cross validation of the male group. The participants were labelled 'P01', 'P02', etc. The number indicates the participants' collection order. In each round, one subset was chosen to be testing data samples, and others were training sets.

son, 2000). Therefore, the whole dataset was separated into two groups. One group included 44 females; another group included 29 males.

- Eye states: Our EEG data was recorded in two eye states: eyes open and eyes closed. Each eye state lasted 1 minute. Since each data sample would be split into equilong epochs, each epoch only included eyes open or eyes closed data. Therefore, we did not need to do special processing in this stage.

3.3.2 Feature Extraction & Predictive Analysis

The feature extraction and predictive analysis methods are the most important parts of depressivity prediction models. Feature extraction is the step of extracting the features in raw EEG data and reducing the signal dimension to make the feature fit the prediction methods. The predictive analysis methods are used to classify or regress the features. The details of these methods are introduced in Chapter 5 and 6. The different combinations of feature extraction and predictive analysis methods formed different depressivity prediction models.

3.3.3 Depressivity Estimation

The next step is the depressivity estimation.

In classification-based models In classification-based models, each epoch of each testing data sample was classified into a high or low class. Because we needed the depressivity of each participant, the depressivity of participants was then estimated by calculating the ratio of high class epochs in all epochs. The depressivity score obtained was called the machine learning depressivity (ML-d) score, and in the classification models, ML-d was calculated by Equation 3.2:

$$ML = \frac{number\ of\ 'high'\ epochs}{number\ of\ all\ epochs} \times 100\% \tag{3.2}$$

For example, there were 230 epochs of data sample 'P03' for testing. The classification model predicted 190 epochs as high class, and 40 epochs as low class. Then the ML-d score of 'P03' was 82.61%.

As mentioned in Section 3.3.1.1, in each experiment, A and B data samples were tested only once and obtained a ML-d score. However, intermediate samples (Sample No. 13-61 in Figure 3.4) were all tested twice. Therefore, the ML-d scores of intermediate samples in each experiment were the average of two rounds.

In regression-based models In regression-based models, each epoch of each testing data sample received one prediction value. The ML-d score of each testing participant was estimated by averaging all epochs' prediction value:

$$ML = \frac{\sum_{i=1}^{n} P_i}{n} \tag{3.3}$$

where n is the number of data sample's epochs. P_i is the prediction value of ith epoch.

3.3.4 Model Performance Evaluation

We tested the effectiveness of the depressivity prediction model by correlating the PID5-d score against the ML-d score.

Pearson's R is the value of the linear correlation coefficient and measures the strength and the direction of a linear relationship between two variables (Fisher, 1992). The formula for R is (The MathWorks, Inc., 2018a):

$$R = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{3.4}$$

where $cov(X, Y)$ is the covariance of X and Y. σ_X is the standard deviation of X, σ_Y is the standard deviation of Y.

The range of R value is $-1 \leq R \leq +1$. If R is close to $+1$, it shows that two sets of data have a strong positive linear correlation. If R is close to -1, it shows that two sets of data have a strong negative linear correlation.

We tested the correlation for significance and, for the purposes of numerical comparison between methods, converted its value to percent variance (i.e., $R^2 \times 100$). We also calculated the corresponding p-value for testing the hypothesis of no correlation. The P-value is the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero (Romero, 2013). In this thesis, we set the significance value as 0.05 which is the most typical setting. If a p-value is smaller than the 0.05, then the corresponding correlation is considered significant.

3.4 Data Preprocessing

Since artefacts in EEG signals may affect signals of interest, the raw EEG data needed to be preprocessed to improve the signal's quality. In our study, we used six steps to remove the artefacts. They combined different methods: baseline drift removal, mains noise removal, eye blink removal, data splitting, and rejecting extremely contaminated epochs automatically and manually (Shown in Figure 3.6).

3.4.1 Baseline Drift Removal

Baseline drift is caused by various reasons such as temperature, the instrumentation's bias, loose contact of electrodes and so on (Reddy and Narava, 2013). Therefore, the first step of data preprocessing was applying a high-pass filter to remove the baseline drift. Winkler, Debener, Müller, and Tangermann (2015) recommend a 1-2Hz high-pass filter to remove the baseline drift. To retain EEG information as much as we can, in this project, a 1 Hz FIR high-pass filter was applied to remove baseline drift. Figure 3.7 shows a comparison of an original signal and the same signal removed baseline drift (The original signal was randomly picked from our dataset.). The figure shows that after 1 Hz high-pass filter, the baseline drift was removed.

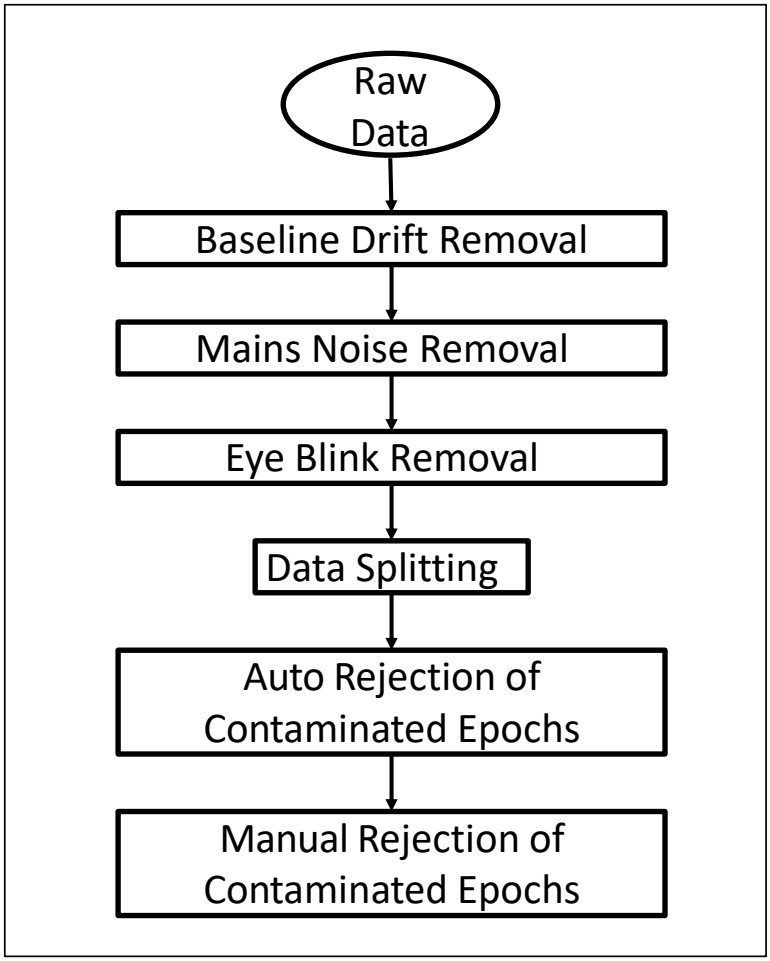

Figure 3.6: This figure shows the data preprocessing steps.

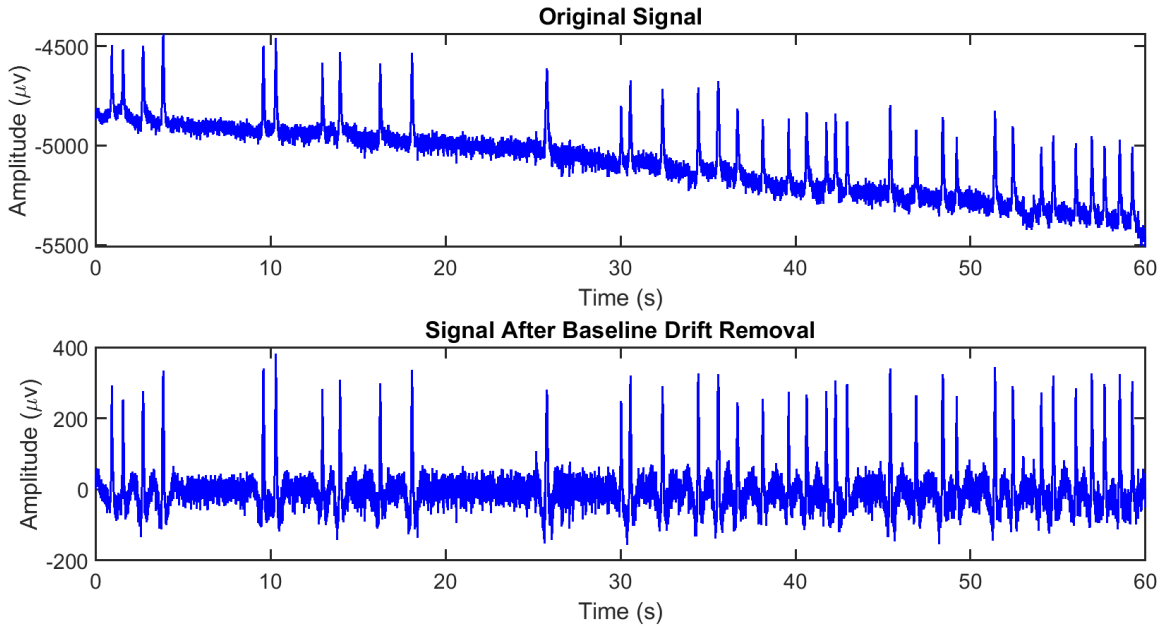

Figure 3.7: This figure shows a comparison of original signal and the signal removed baseline drift. The figure shows that after 1 Hz high-pass filter, the baseline drift was removed.

3.4.2 Mains Noise Removal

EEG signals are contaminated by mains power from electrical equipment. The mains power causes a local peak in the EEG signal around the frequency of electricity. Therefore, the second step of data preprocessing was to remove the mains noise. According to the electricity (safety) regulations (New Zealand Parliamentary Counsel Office, 2010), the frequency of electricity supplied must be maintained between 47 Hz and 53 Hz. In our dataset, the artificial spike caused by mains power is from 49.9 Hz to 50.1 Hz (see left of Figure 3.8).

Our dataset recorded EEG data from several channels synchronously, therefore, mains power should have a fixed phase all through the record and be synchronous across all channels whereas real brain signals are not the same phase as each other and should not be synchronous. Therefore, to remove the mains noise, a phase-fixed method was designed to evaluate the component of mains noise in EEG signal.

- Step 1: Determining the frequency range which suffered from mains noise. In our dataset, each data sample was inspected and the results show that the artificial spike caused by mains power is from 49.9 Hz to 50.1 Hz (left side of Figure 3.8).

In order not to miss any mains noise component, the range was set from 49.8 Hz to 50.2 Hz.

- Step 2: Averaging signals of all (18) channels to reduce the real brain signals and enhance the mains noise.

$$M = \frac{\sum_{i=1}^{n}(O_i)}{n} \tag{3.5}$$

where, n is the number of channels (n = 18). O_i denotes the original signal of the *ith* channel. Because the real EEG signal has different phases and mains noise has the same phase in different channels, summing all signals can retain the mains noise while reducing the real EEG signal.

- Step 3: Using FFT algorithm to calculate the amplitudes and phases from 49.8 Hz to 50.2 Hz of signal M (obtained by Equation 3.5).

- Step 4: Using the amplitudes and phases which were obtained in Step 3 to generate sine waves to simulate the mains noise component of the data sample and scaling the sine waves back to each channel. This step was repeated from 49.8 Hz to 50.2 Hz. The length of whole signal is around 122880 and the sample rate is 256 Hz. Therefore, the frequency interval is around 0.002 Hz and the number of iteration is around 200.

  In the first iteration, the amplitude and phase of 49.8 Hz of signal M was used to generate a 49.8 Hz sine wave T. The mains noise have a fixed phase across all channels while changing its size. Therefore, the sine wave T was scaled for each channel using the slope estimated by robust regression between the sine wave T values and the EEG values for that channel. The scaled values were then subtracted from the original EEG to leave EEG free of 49.8 Hz mains noise. The procedure was repeated for each frequency step up to 50.2 Hz.

  In the procedure, robust regression with a bisquare weight function was used to scale the sine wave T to each channel, because compared to ordinary least squares, robust regression can reduce the influence of outlines of original signal.

After these steps, the mains noise of each channel was removed, and the natural EEG was left intact. Figure 3.8 shows a comparison of the raw signal (after baseline drift removal) and the signal after removal of mains noise in the frequency domain. The raw signal was randomly picked from one channel of one participant.

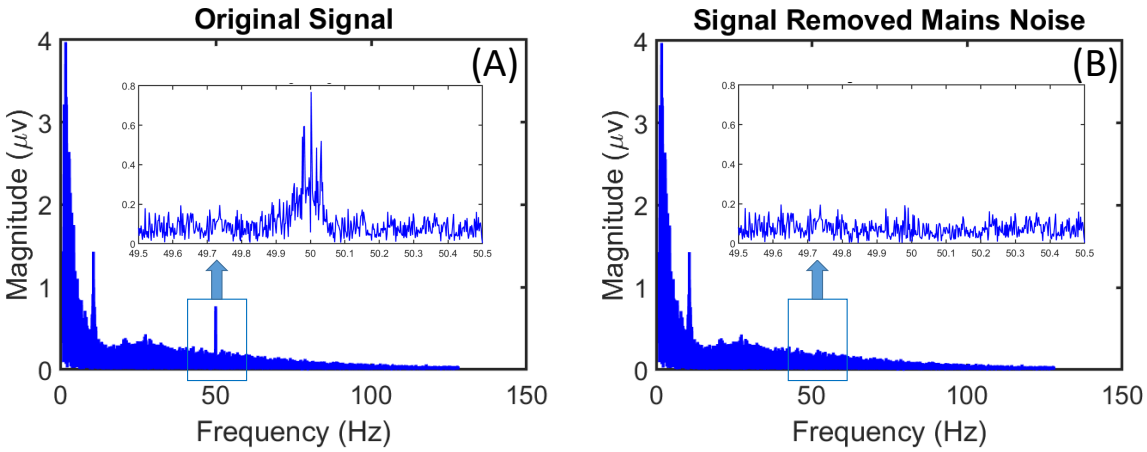

Figure 3.8: This figure shows a comparison of the raw signal (after baseline drift removal) and the signal removed mains power noise in the frequency domain. The raw signal shown was randomly picked from one channel of one participant.

3.4.3 Eye Blink Removal

The EEG signal is always contaminated by eye blinks (Woestenburg, Verbaten, and Slangen, 1983; Gratton, 1998). Therefore, we designed a single-channel physiology-based method to remove eye blink artefact components (see Chapter 4). Figure 3.9 shows an example of an EEG signal (after baseline drift and mains noise removal) before and after removal of the eye blink component.

3.4.4 Data Splitting (Open/Closed)

In the relaxation test, the eye state was varied under instruction from the experimenter in the order 'O-C-O-C-C-O-C-O' (O: eyes open, C: eyes closed). Each eye state lasted 1 minute. EEG recording is influenced by eye state (Rösler and Suendermann, 2013), so the eyes open EEG data and eyes closed EEG data were extracted separately. Meanwhile, in consideration of the duration of the change in eye state, the first and last second of each eye state EEG recording were discarded. Because each eye state has 4 minutes ($4 \times 60 = 240$ seconds) recording, after discarding, there were 232 seconds of EEG recording for each eye state left.

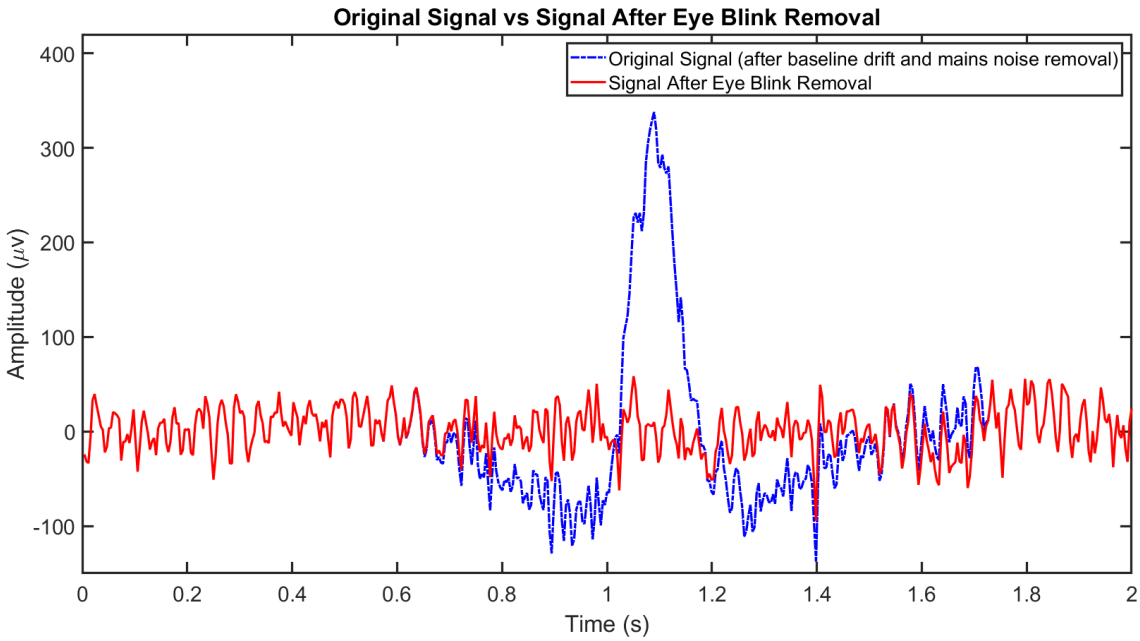

Figure 3.9: This figure shows a comparison of the original signal (after baseline drift and mains noise removal) and the signal after eye blink component removal. The blue dotted line shows the signal which was after baseline drift and mains noise removal and the eye blinks remained. The red solid line was the clean data, and it shows the eye blink was removed.

3.4.5 Auto Rejection of Contaminated Epochs

Because of head movement and other interference, sometimes extremely high voltage values are recorded. The eye blink removal method cannot deal with this kind of unpredictable artefact, so an amplitude threshold (-500 to 500 μV) was set to detect this kind of artefact.

Each EEG data sample was split into 1-second length epochs. If the epoch included values which exceeded the threshold, the epoch was rejected.

3.4.6 Manual Rejection of Contaminated Epochs

After all automatic processing, each data sample was visually inspected, and some contaminated epochs that had been missed by the algorithms were rejected. Britton, Frey, Hopp, Korb, Koubeissi, Lievens, Pestana-Knight, and St Louis (2016) listed the main types of artefacts of EEG and the records were visually searched for this and then discarded.

After that, the clean dataset was ready to be used in depressivity prediction models.

3.5 Conclusion

This chapter provides an overview of the design of the depressivity prediction models. It first introduced the EEG dataset that was used to design the models. The introduction includes the collection procedure of the dataset, the distribution of the PID-5 depressivity scores, and the procedure used to tidy the dataset.

Next, it provided an overview of depressivity prediction models. The models include five basic steps: the participant assignment, the feature extraction, the predictive analysis, the depressivity estimation, and the model performance evaluation.

Finally, it introduced the data preprocessing procedure that we used to provide a useable clean dataset.

Chapter 4

Eye Blink Removal

EEG signals are often contaminated with artefacts, particularly with large signals generated by eye blinks. Deletion of artefacts can lose valuable data. Current methods of removing the eye blink component to leave residual EEG, such as blind source component removal, require multichannel recording, are computationally intensive, and can alter the original EEG signal.

Chapter 3 introduces the data preprocessing methods used in this research, the step of eye blink artefact removal is implemented by a novel single-channel physiology-based eye blink artefacts removal method. In this chapter, the method is described. In the rest of this thesis, we refer to it as the template eye blink removal method (TEBR) since the key step of this method is to construct a template of each eye blink.

TEBR removes the blink component, leaving uncontaminated EEG largely unchanged. Processing time allows its use in real-time applications such as neurofeedback training.

Blink removal had a success rate of over 90% recovered variance of original EEG when removing synthesised eye blink components. Fronto-lateral sites were poorer (~80%) than most other sites (92-96%), with poor fronto-polar results (67%).

When compared with three popular independent component analysis (ICA) methods, TEBR was only slightly (1%) better at frontal midline sites but significantly (>20%) better at lateral sites with an overall advantage of ~10%.

With few recording channels and real-time processing, TEBR shows clear advantages over ICA for removing eye blinks. It should be particularly suited for use in portable brain-computer-interfaces and in neurofeedback training.

The content of this chapter has been published in *Journal Neuroscience Methods* (Zhang *et al.*, 2017).

4.1 Introduction

Brain signals in the human scalp EEG can be heavily contaminated with non-neural artefacts. In particular, eye blinks produce very large potentials and can occur at a rate of about 20 times per minute, lasting between 50 to 500 milliseconds (Halder, Bensch, Mellinger, Bogdan, Kubler, Birbaumer, and Rosenstiel, 2007).

Here, we propose an approach to artefact removal that is based on the physiology of the eye blink and is similar to Morphological Component Analysis (MCA) but does not require any prior knowledge of the precise shapes of individual eye blink artefacts in the EEG. It uses a generic skeleton of an eye blink artefact, based on the ballistic properties of eye blink generation (described in more detail in the methods), which is then fitted to each individual eye blink on a single channel (such as Fp1, which is close to the eyes) to create a template that, as in MCA, is then used to remove that particular eye blink waveform from all channels (including the channel used to derive the template) leaving residual EEG. In the results we compare this method with ICA+ADJUST (Automatic EEG artifact Detection based on the Joint Use of Spatial and Temporal features) applied to the same data. Given the adaptive nature of the model used, this removal is highly specific to the eye blink component of an individual. Importantly, it only alters the EEG in the region at which an eye blink is detected. This single channel processing, and multichannel removal, has a low computational overhead and so can be carried out on the fly when the EEG data are processed, and, so it could be used for real-time eye blink removal in neurofeedback and brain-computer-interface studies.

4.2 Method and Materials

4.2.1 Participants & EEG Recording

Determination of eye blink parameters of the key population and of the algorithm's effectiveness were carried out on data obtained from 64 participants (46 females, 16 male; aged 18-37 years, mean 21.87 years). EEG was recorded from electrode sites F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8 and Fp1 (International 10-20 system). The data analysed in the current chapter is different from the dataset introduced in Section 3.2 of Chapter 3. The data samples used here had previously been band pass filtered (1-36 Hz) and the original sampling rate of 256Hz had been down-sampled for analysis to 128 Hz. The purpose of down-sampling was to reduce

file size and for compatibility with previous datasets collected for other projects. So all examples were based on 128Hz sample rate. Other recording and participants details were the same as introduced in Chapter 3.

4.2.2 Eye Blink Removal Method

TEBR method consists of two key steps. First, a generic skeleton of an eye blink artefact, made up of straight lines, is fitted to an individual eye blink detected automatically (as detailed below) in the EEG recorded from Fp1. The intersections of the straight lines are then progressively smoothed to a best fit with the local EEG. Next, the entire fitted eye blink is scaled to the current channel with a conventional least squares technique (Gratton, 1998) and the result is subtracted to leave residual EEG. Scaling and subtraction is repeated for each channel (including Fp1). The whole procedure is then repeated for the next eye blink.

Below, we present the theoretical basis for the design of the generic skeleton, followed by detailed descriptions of the algorithm.

4.2.2.1 Physiological basis of the eye blink artefact skeleton

Our generic skeleton of an eye blink artefact is based on a specific physiological property of eye blinks – they are ballistic movements (Evinger, Shaw, Peck, Manning, and Baker, 1984). That is, like a bullet, their trajectory and amplitude is determined by the size of an essentially single impulse. In terms of lid position, the eye blink involves a single fast downward component and a slower return component – a simple U-shaped curve. Lid velocity (the first differential of lid position) has an extra component generating a curve with two inflections (~), which can have a slight inflection at the intermediate zero (baseline) point (Evinger *et al.*, 1984). The electrical artefact generated by the eye blink is reported to essentially follow the form of the velocity curve (not lid position) but can have an additional late recovery component as well.

4.2.2.2 Modelling the components of an eye blink artefact with straight lines

The critical thing for our analysis is that primary deflections involved in all the components of the electrical artefact generated by the eye blink can be treated as essentially straight lines and the change-over points can be closely modelled by simple smoothing of the intersection point of the lines. Figure 4.1 shows a standard eye blink, which

is composed of three main U-shaped components followed by a subsidiary, smaller U-shaped component resulting in 4 inflection points. The components occur in the order negative-positive-negative-positive (relative to a zero baseline) or vice versa (positive-negative-positive-negative), depending on the position on the scalp.

We start with detection of a potential eye blink peak (Inflection 2 in Figure 4.1) in Fp1 by the crossing of a positive threshold (determined separately for each individual in the current version of the algorithm). The preceding opposite inflection (Inflection 1 in Figure 4.1) is then detected as a local minimum. A straight line is fitted via conventional regression to the centre portion of the rising phase of the blink between these two inflections and a second straight line fitted to the centre of the initial downward portion of the blink. These two lines are then extrapolated until they intersect and are subjected to successive 3-point smoothing (which progressively affects points further from the original intersection) until the template in the inflection region is at its best fit to the original data (i.e. until the correlation of the raw data and model data ceases to increase). Each of the remaining inflections is similarly modelled with two straight lines (the two sides of the U-shape) and all line intersections are smoothed – including where the initial and final model lines meet the zero baseline. Note that the baseline for the template must have zero values to prevent it altering the EEG during correction. The resultant template can then be used to correct all channels, including the channel used for its construction. A variety of rules applied to, e.g., the expected location of the inflections prevent the template from being fitted to artefacts that do not have the typical shape of an eye blink.

4.2.2.3 Locating inflections

Of the standard recording sites in a typical EEG system, Fp1 (or Fp2 or Fpz) is closest to the eye and most affected by eye blinks. Fp1, therefore, has the greatest eye blink signal to noise ratio (and vice versa for EEG signals). So we used Fp1 as the source channel from which to create our blink templates. Over the time course of a recording session, participants produce blinks that can vary substantially in size. Variations in shape also occur. We therefore constructed a separate template for each blink detected in Fp1. Basic detection of a blink used a simple thresholded window. We first determined the maximum positive value less than $500\mu V$ (to exclude non-ocular artefacts) in the entire Fp1 record. We then set the blink detection window to the range 25-100% of the peak value of the largest blink. Note that for real-time use of the procedure an initial 'calibration' recording period could be used (to locate the peak value), which could

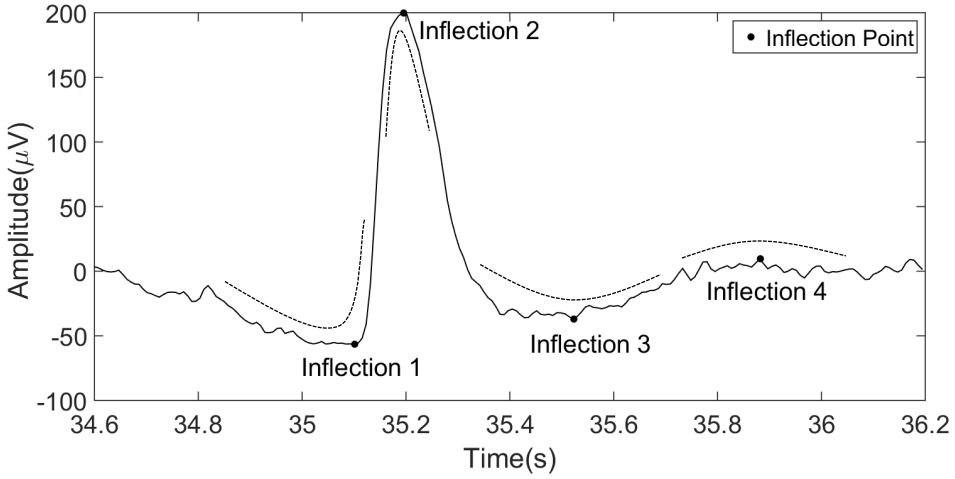

Figure 4.1: **An example of an eye blink from a participant.** The dotted U-shaped curves show the three main components (centred on inflections 1, 2 and 3) of a standard eye blink. These are very often also followed by a fourth component (centred on inflection 4), which is negligible in this example.

include periods when the participant was requested to blink strongly. The Fp1 record was then scanned and the first EEG value falling within the window was taken as the 'Start' point (see Figure 4.2) for construction of the first blink template. Given we are using Fp1 and with the window set to positive values (relative to baseline), this start point will generally be in the rising phase of the artefact somewhat before inflection 2. After construction of the blink template, and removal of the blink components, the next EEG value that falls in the window is taken as the 'Start point' for the next template and the whole procedure is repeated until the end of the recording is reached.

4.2.2.4 Constructing the template

Given the 'Start point', we first estimated the positions of the 'inflection' points (Figure 4.1) and then the 'zero' points (where the record returns to baseline). The rules for determining these estimated points are given in Table 4.1, the results are shown in Figure 4.2a. The estimated inflection and zero points were then used to locate 'ends' (start and finish points) for positioning the straight lines that define the main skeleton of the template. The rules for determining the ends of the lines are given in Table 4.2 and the results in Figure 4.2b.

Each eye blink template was built from 9 straight lines (Figure 4.2). Line 0 and

Inflection	thing; from : to	Zero	thing; from : to
I1	MIN; S-500ms:S	Z1	~ZERO; I1-500ms:I1
I2	MAX; I1:I1+500ms	Z2	~ZERO; I2:I3
I3	MIN; I2:I2+500ms	Z3	~ZERO; I3:I4
I4	MAX; I3:I3+500ms	Z4	~ZERO; I4:I4+390ms

Table 4.1: **The detection rules for finding nominal 'inflection' points and 'zero' points.**The detection rules for finding nominal 'inflection' points and 'zero' points. The same values are used for all blinks and all participants. These are found by a search for a particular thing – a local maximum (MAX) or minimum (MIN) or the closest data point to the zero value baseline (~ZERO) – through the EEG from one particular time (e.g. S-500ms = 500ms before the Start point) through to a second particular time (e.g. S = Start point; i.e. the point at which the positive going component of the eye blink passes the detection threshold). These inflection and zero points are then used to position 'ends' of lines (see Table 4.2). Note that the value of 390ms reflects our use of the round number of 50 samples at 128Hz, based on initial coarse adjustment of this parameter to optimise detection. If the local baseline, over the period being analysed is not zero (as when a high pass filter is not used) then the data segment being fitted needs to be adjusted to zero baseline before fitting (or the baseline value used to detect the ZERO points and the template adjusted after fitting); i.e. the first and last line segments of the template should consist of zero values before the template is used for corrections.

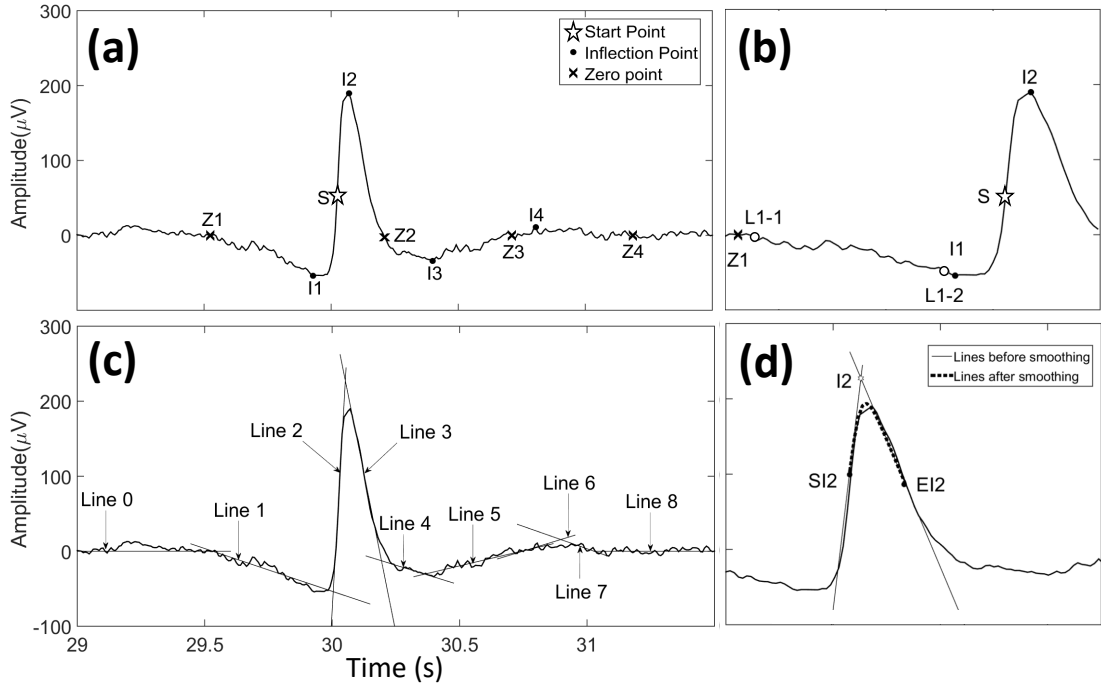

Figure 4.2: **Illustrations of 'Start point', 'Inflection point', 'Zero point', lines, template construction, and smoothing.** (a) The 'Start point' (☆) is the first EEG value that falls within the 25-100 % of maximum peak range for that participant. Inflections (e.g.• I1) are local extremums. Zero points (e.g. × Z1) are the points at or just prior to when the signal crosses the zero (average baseline) value. (b) Template construction starts with determining the end points defined in Table 4.2 and the fitting of least squares linear regression lines to the EEG between them. For example, Line 1 was fitted between end points L1-1 and L1-2. (c) 9 straight lines were used with Line 0 and Line 8 being two straight lines at amplitude 0 V and parallel to the time axis. (d) The intersection of each pair of lines was then subjected to repetitive 3-point smoothing to achieve a least squares best fit to the EEG.

Line 8 are fitted, with zero slope, to the baseline EEG before and after genuine eye blink components. Lines 1 through 7 represent eye blink components. The line 6 and line 7 components can be small (see example in Figure 4.2), or essentially absent, as the late phase of the blink is very variable. Each straight line was generated using the least square method to best fit the EEG between the pair of ends defined in Table 4.2. For example, 'Line 1' was obtained by fitting the data between 'L1-1' and 'L1-2'. Once they had been fitted, adjacent lines were extrapolated until they intersected.

Ends	Position	Ends	Position	Ends	Position
L0-1	Z1-625ms	L3-1	10% I2:Z2	L6-1	10% Z3:I4
L0-2	Z1-78ms	L3-2	90% I2:Z2	L6-2	90% Z3:I4
L1-1	10% Z1:I1	L4-1	10% Z2:I3	L7-1	10% I4:Z4
L1-2	90% Z1:I1	L4-2	90% Z2:I3	L7-2	90% I4:Z4
L2-1	10% I1:I2	L5-1	10% I3:Z3	L8-1	Z4+78ms
L2-2	90% I1:I2	L5-2	90% I3:Z3	L8-2	Z4+625ms

Table 4.2: **The detection rules for 'ends' of lines.** The detection rules for 'ends' of lines (L0:L8), used to generate straight lines of the eye blink templates. I1-I4, Z1-Z4 are as in Table 4.1. Most line ends are positioned at 10% of the distance between the relevant I or Z point. L0 and L8 have line ends that are not located on the existing EEG but are located at the Y-axis origin (i.e. they represent zero correction of the EEG) and so L0 and L8 are both parallel to the X-axis. The value of 625ms reflects our use of the round number of 80 data points in our algorithm, and 78ms reflects our use of 10 data points, at 128Hz.

The intersection zones were then repetitively three-point (average of current point with previous and following points) smoothed while the correlation coefficient with the original data continued to increase. The smoothing started at the mid-point of the first line (e.g. SI2 in Figure 4.2) and continued to the midpoint of the second line (EI2). Note that 3 point smoothing does not affect the values of the straight portions of the lines and so, on the first iteration, only the peak (intersection point) is smoothed, on the second iteration only the peak and the two adjacent data points are smoothed and so on.

4.2.2.5 Template adjustment

There are a number of conditions where simple application of the basic algorithm will give incorrect results. In these cases, tests are applied for the relevant condition and a suitable adjustment is made to the basic model.

The Crossover of Line 2 and Line 3 should normally be above Inflection 2 (as in Figure 4.2). If it is not (as in Figure 4.3a), after smoothing, the peak of the model would be far away from original eye blink's peak. So, when the amplitude of Crossover 2 is less than the amplitude of Inflection 2, the percentage positions of L2-1 and L2-2 (10%

and 90% respectively at the start) are brought closer in steps of 1% until Crossover 2 is above Inflection 2 (see Figure 4.3b) or their values are 30% and 70% respectively (beyond which point the fitted line would not be able to accurately reflect the ballistic component of the eye blink). The normal procedure then continues.

As mentioned above, a standard eye blink has three main components and they are usually followed by a fourth component. But a few blinks do not fit the basic rules. The fourth component and, in some cases, even the third component may be missing. In these cases, the template model was adjusted as follows using criteria developed during initial experimental adjustments of the algorithm to optimise detection.

1. The most common atypical eye blink lacks the fourth component. There are three indicators of this kind of eye blink: 1) the interval between points L5-2 and L6-1 exceeds 180 ms (see Figure 4.3c); 2) Inflection 4 and Zero 4 coincide (see Figure 4.3d) – this is usually because two blinks are very close; and 3) Line 7 is parallel to Line 8 – which occurs when the EEG shows no inflection matching a fourth peak (this also always happens when two blinks are close).

2. Less frequently, atypical eye blinks have only two components and lack the third and fourth components. There are two indicators of this kind of eye blink: 1) the interval between points L3-2 and L4-1 exceeds 125ms (see Figure 4.3e); 2) As shown in Figure 4.3f, Intersection 4 (the intersection of Line 4 and Line 5) is above Intersection 3 (the intersection of Line 3 and Line 4). In either of these cases, we generate our template only using 5 lines (Line 0, Line 1, Line 2, Line 3, and Line 8).

4.3 Results

4.3.1 Removing Blinks

The eye blink potential propagates across channels retaining its shape while changing its size. We therefore scaled the template for each channel using the slope of the linear regression between the template values and the EEG values for that channel (so the regression slope for Fp1 was 1). The scaled values (which represent deviations from the baseline) were then subtracted from the original EEG to leave artefact free EEG (Gratton, 1998).

Figure 4.4 shows EEG from a participant before (dotted line) and after (solid line) application of the blink removal procedure. There is a typical eye blink artefact in the

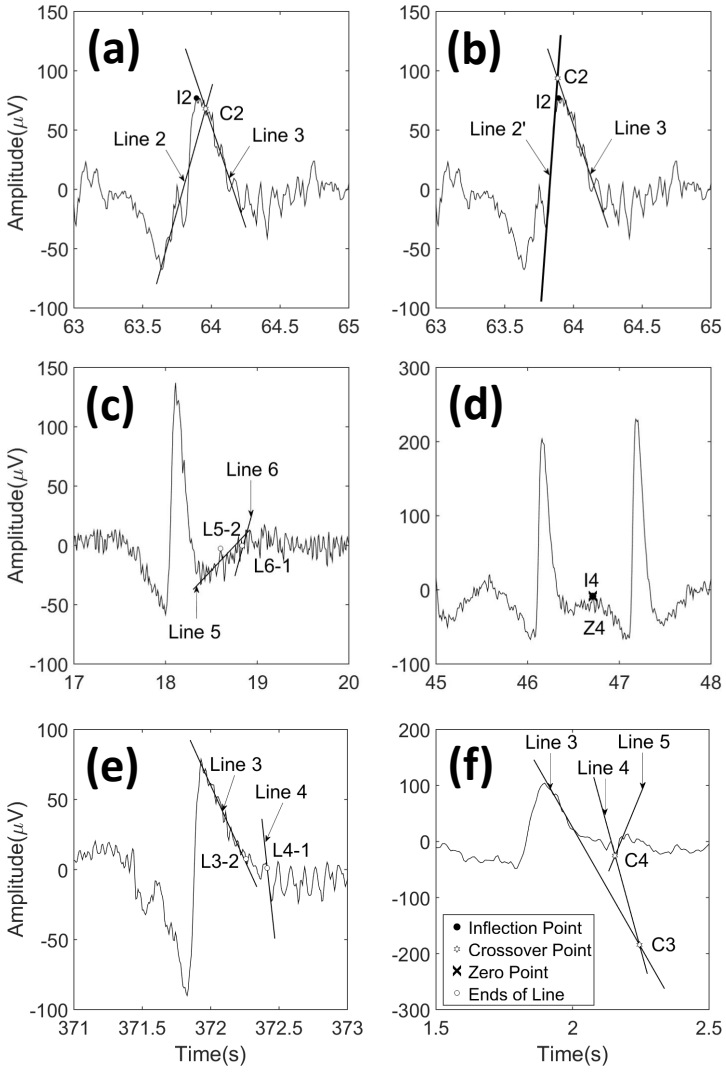

Figure 4.3: **Template adjustments.** If Crossover 2 (the intersection of Line 2 and Line 3) is below Inflection 2, after smoothing, the peak of the model will be far away from the peak of the original eye blink (a). In this situation, we adjust the positions of L2-1 and L2-2 in 1% steps towards each other until, (b), Crossover 2 is above Inflection 2 or L2-1 (for further details see text). Some eye blinks lack the fourth component. (c) shows one indicator of this – L5-2 and L6-1 are separated by more than 180 ms. (d) shows a second indicator: Inflection 4 and Zero 4 coincide. Some eye blinks lack both the third and fourth components. (e) shows one indicator: L3-2 and L4-1 are separated by more than 125ms. (f) shows a second indicator: Crossover 4 is above Crossover 3. For (e) or (f), we use only 5 lines in the model (Line 0, Line 1, Line 2, Line 3, and Line 8).

interval of 4 s to 5.6 s and it can be seen that almost all channels' data are influenced by the eye blink.

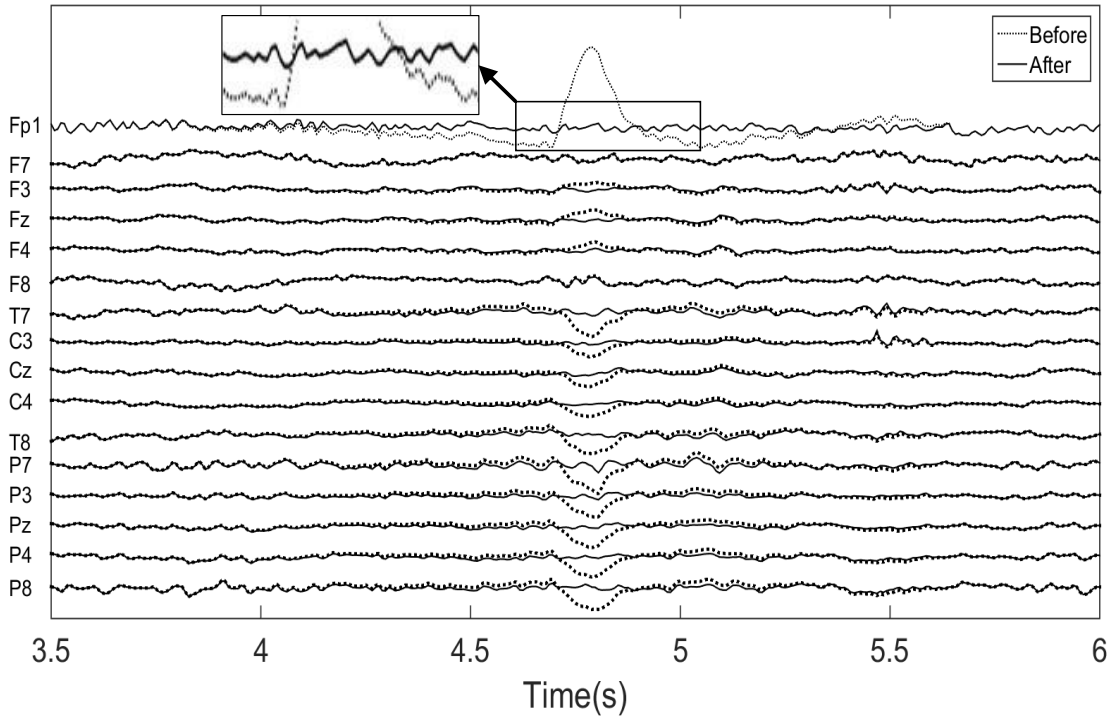

Figure 4.4: This figure shows a comparison of data before (dotted lines) and after (solid lines) eye blink removal. Blinks were detected and the template created using Fp1 and then eye blinks removed across all 16 channels using the same template scaled via the slope of a least squares regression line (Gratton, 1998). There is a typical eye blink artefact in the interval of 4 s to 5.6 s and it can be seen that almost all channels' data are influenced by the eye blink. After removal (inset) ongoing rhythmicity becomes clear.

4.3.2 Validation

4.3.2.1 Blink artefact simulation

Successful eye blink removal on a percentage basis was assessed over all 64 participants. To assess the detailed effectiveness of TEBR procedure in terms of recovering the

original uncontaminated EEG, we randomly picked 5 participants and added simulated eye blink wave forms to existing artefact-free EEG and then tested the capacity of TEBR method (and of ICA) to recover the original EEG.

For each participant, we visually inspected their EEG and extracted the first 100 segments of EEG data (each segment is 200 samples/1.57s long) contaminated with eye blink artefacts. For participants with fewer than 100 contaminated segments, we reused part of the initial set of blinks to make the total up to 100, i.e. we reused the first 10 segments for a participant with only 90 eye blink contaminated segments. Next, we aligned the peak of the blink in each segment (70 samples/547ms before the peak and 130 samples/1016ms after the peak). These were then averaged in sets of 10. The average was then low pass filtered (14 Hz) and smoothed with a 3-point running mean. Both the averaging and smoothing were intended to retain the scalp distribution of the blinks while eliminating background EEG. This procedure resulted in 10 synthetic eye blinks for each participant. The synthetic blink deviation components were added to the raw EEG starting after the 100th data point (781ms) of a 3s segment of artefact-free EEG taken from of a period of at least 3 s without artefacts; and with the 10 artefact-free segments required spread out within the 8-minutes of recording. Figure 4.5a shows an example of a simulated eye blink at low amplification. Figure 4.5b shows the segment of artefact-free EEG, with and without the addition of the eye blink shown in Figure 4.5a, at higher amplification.

4.3.2.2 Comparison of TEBR method with ICA

We used three popular ICA algorithms (SOBI (Belouchrani, Abed-Meraim, Cardoso, and Moulines, 1997), Infomax (Bell and Sejnowski, 1995) and FastICA (Hyvärinen and Oja, 2000)) with ADJUST method to compare with TEBR method. The main idea of ICA is to decompose a multivariate signal into statistically independent non-Gaussian components (Hyvärinen, 2013). Each row of the original EEG matrix is the summation values originating from different sources. The ICA method transfers this original matrix into a new matrix, where each row of the new matrix contains values estimated as being from one independent source. These sources represent different activities originating from one or more cortical patches, and from different types of artefacts, e.g., eye blinks, eye movements, or muscle activity (Arnaud, Delorme and Scott, Makeig, 2018). ICA can be performed by different algorithms. The three ICA algorithms what we used are commonly used in brain signal separation (Sahonero and Calderon, 2017; Congedo *et al.*, 2008). An ICA algorithm can separate different independent sources. However,

its implementation is semi-automatic. It requires a trained person to identify the specific eye blink component by visual inspection. This process is time-consuming and user-dependent. An automatic algorithm, ADJUST (Mognon *et al.*, 2011), has been developed to automate the manual step. This algorithm was used in the current thesis. Overall, ICA+ADJUST methods derive and delete the estimated independent noise components from the entire signal, while our TEBR method first locates the eye blink area and then removes the specific eye blink component from only the eye blink area.

All these methods were applied to the entire 8 minutes of data from each participant and then, after reconstitution of the signal, the relevant 1.57s segments to which the blinks had previously been added were extracted for analysis. Figure 4.5c shows the data after application of our TEBR method. Figure 4.5d-f show the results with processing by Infomax, SOBI and FastICA, respectively, for comparison. In all of Figure 4.5b-f the data of interest are plotted as solid lines overlaid on the original uncontaminated data plotted as dotted lines. In this example, TEBR, SOBI, and FastICA, but not Infomax removed the eye blink artefacts. The problem with Infomax relates to instabilities in ADJUST that can affect all three ICA methods and that can be dealt with by inspecting all components visually and deleting the artefacts that ADJUST missed. Importantly, the two successful ICA methods also altered some original brain wave information in non-eye blink areas (e.g. Figure 4.5e,f at 2.5-3s, particularly in Fp1) as has been reported previously (Inuso *et al.*, 2007; Pontifex, Gwizdala, Parks, Billinger, and Brunner, 2017). Note that given the restricted operation of our eye blink template, it retains completely unchanged values in areas where there is no eye blink.

To quantify the degree of fit of the processed data to the original clean EEG, we calculated the correlation coefficient between the raw and cleaned data for each method for the 5 participants. After application of the relevant eye blink removal procedure, the raw:cleaned correlation was calculated for each blink segment and then averaged across the 10 segments, excluding those where ICA failed to remove blinks. Correlations were converted to percent of shared variance ($r^2 \times 100$) (shown in Table 4.3) and submitted to two ANOVAs: one over all 16 channels and one excluding Fp1 to allow a 3 (anterior-posterior) × 5 (medial-lateral) trend analysis. The histogram on the right of Figure 4.6 shows that the mean percent covariance is highest for the TEBR method (91%) and lower (78-82%) for the ICA methods. ICA correction was generally good at posterior and midline sites and poorest at antero-lateral sites (Figure 4.6). ANOVA including all correction methods and channels found significant differences between the methods (method F (1.742, 6.969) = 8.640, p = 0.014 with Greenhouse-Geisser correction of df;

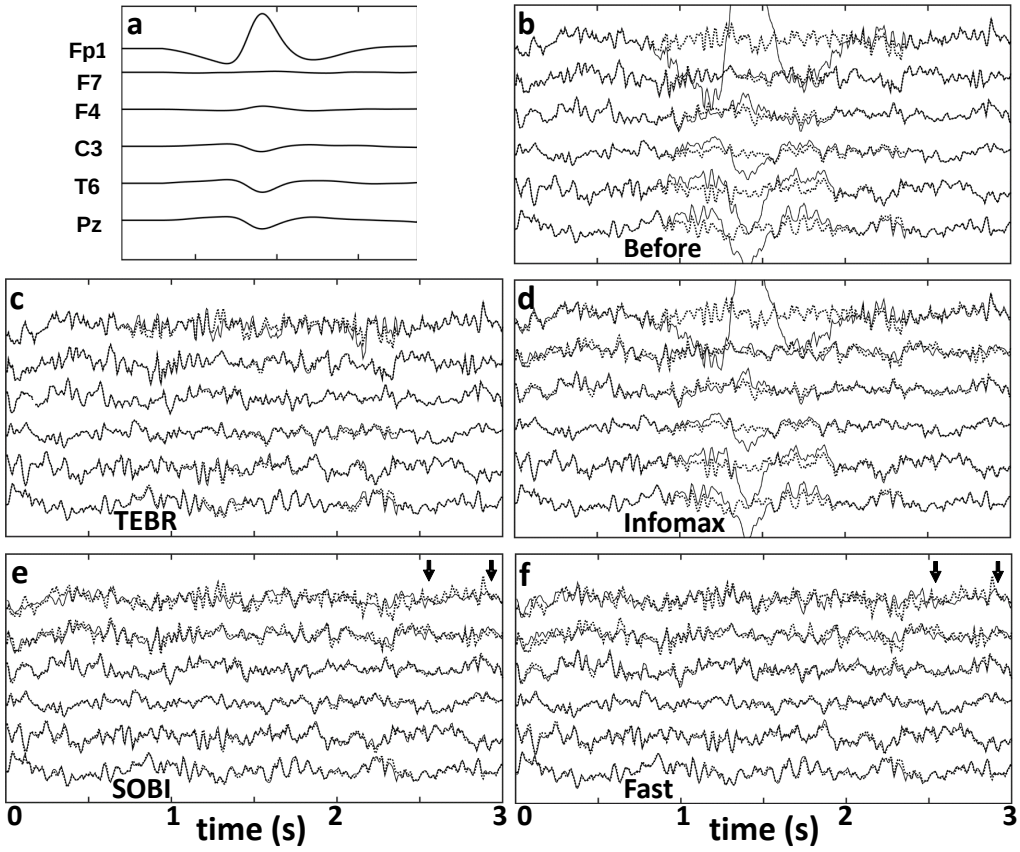

Figure 4.5: **Synthesised eye blink and testing** Original uncontaminated EEG is shown as a dotted trace with solid traces for the relevant modified data. (a) Example of synthesised eye blink obtained by averaging 10 blinks from one participant and then filtering with a low pass filter (14 Hz) and smoothing with 3-point smoothing. The 200 sample/1.57s long segment is located relative to the two panels below so it is located immediately above the point in them at which it was inserted prior to their attempt to remove it. This procedure retains the conformation and size of the eye blink across the 16 channels. (b) Uncontaminated EEG before and after addition of the synthesised blink component shown in (a). (c) Comparison of the uncontaminated EEG with the results of TEBR procedure. (d) As (c) but using Infomax ICA + ADJUST instead of TEBR procedure. (e) As (d) but using SOBI ICA + ADJUST. (f) As (d) but using FastICA + ADJUST. Vertical arrows mark the start and end of a section where ICA that successfully removed eye blinks altered artefact-free segments of the EEG.

method × channel F (2.871, 11.499) = 2.832, p = 0.087 Greenhouse-Geisser/p<0.001 uncorrected). Analysis excluding Fp1 and analysing for anterior-posterior and medial-lateral trends showed that the tendency for poor correction at antero-lateral sites was

Channel	TEBR	Sobi	Infomax	FastICA
F7	79.86%	46.31%	48.08%	42.54%
F3	94.64%	88.21%	83.44%	76.59%
Fz	95.56%	94.65%	87.68%	94.85%
F4	92.51%	87.79%	87.45%	86.08%
F8	81.49%	55.70%	59.47%	48.19%
T7	94.67%	90.60%	74.91%	75.66%
C3	93.17%	89.04%	89.82%	81.33%
Cz	95.24%	93.97%	94.10%	94.39%
C4	94.14%	88.59%	90.43%	87.63%
T8	94.72%	91.58%	80.59%	84.92%
P7	94.68%	92.92%	92.64%	86.61%
P3	94.20%	91.70%	92.82%	88.55%
Pz	91.88%	90.29%	92.16%	89.19%
P4	92.74%	90.02%	91.37%	90.05%
P8	93.51%	93.12%	92.93%	92.95%
Fp1	66.82%	32.70%	33.97%	25.14%

Table 4.3: Percentage of covariance between the original and corrected signal for the different correction methods.

significant (AP[linear] × ML[quadratic] F (1,4) = 171.835, p <0.001) with the different methods tending to produce different patterns of results (methods F (1.707, 6.829) = 5.441, p = 0.042 Greenhouse-Geisser; method × AP F (2.802, 11.208) = 3.093, p = 0.073 Greenhouse-Geisser). Post-hoc ANOVA with the TEBR method removed retained the AP[linear] × ML[quadratic] contrast (F (1,4) = 81.530, p =0.001) but no longer had any effects associated with method (all F<1 all p>0.430) showing that the original method-related difference was between the TEBR and other methods.

4.4 Discussion

Human EEG contains many complex components and is always contaminated by artefacts, with particularly large components arising from eye blinks. Here we have demonstrated an automatic eye blink artefact removal method (TEBR) based on a novel,

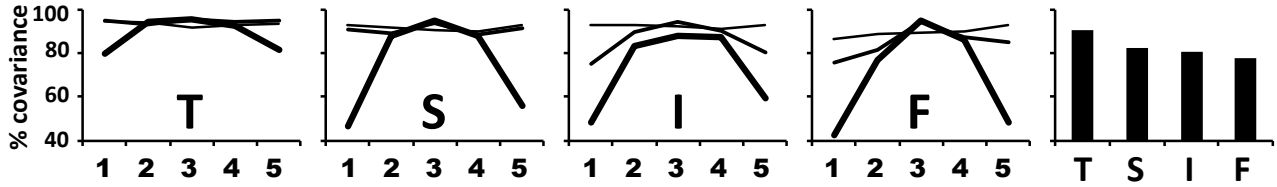

Figure 4.6: Percentage of covariance between the original and corrected signal for the different correction methods. Line graphs are plotted from left (1) to right (5) for each of anterior (thick line), through central (medium line), to posterior (thin line) and exclude Fp1. The histogram shows the average for each method over all channels including Fp1. T = TEBR, S = SOBI, I = Infomax, F = fastICA.

physiologically-inspired, template approach. We have shown that this method can detect and remove eye blink components based on analysis of a single channel and with a success rate of over 90% in terms of recovering the variance of the original EEG prior to addition of synthesised eye blink components.

Blink removal, based on analysis of a single channel with large eye blinks (Fp1) was successful over all channels but with slightly poorer results at the fronto-lateral sites F7 and F8 (~80%) than most other sites (92-96%) and particularly poor results at Fp1 (67%). Fp1 has the largest eye blink component. While this means that Fp1 is the best channel to establish the template in terms of signal to noise ratio it also means that it is where deviations from fit (assuming the same blink morphology across channels) will be greatest. It should be noted that, for this reason Fp1 is also the channel with the poorest correction by ICA (25-34%).

The ballistic movement model can generally deal with double-blinks and triple-blinks (e.g. Figure 4.3d) – however, on occasion the current form of the model fails to fit a complex multiple blink. We have already included some adjustments to reduce the number of remaining failures but further improvements should be possible with additional experimenting. The method would also not be expected to work in cases where the blink was compounded with, e.g., muscle artefact – but these artefacts would remain, in any case, with methods like ICA and would normally be deleted during conventional processing. The method should remove blinks as normal in the presence of epileptiform spikes but, with single channel processing, it might remove some spikes as well (we have not tested if epileptiform data would fit the eye blink

template). Although we have not directly tested this with clinical data, we believe that with whole head multichannel recording, blinks and epileptiform activity could be separated (and the latter preserved) with a simple test of the channel profile of the slope coefficients calculated in the correction step, which would have a mid-frontal loading for eye blinks and a more posterior and likely lateral loading for epileptic spikes. The method might be also extended to other types of artefacts provided their shape could be predefined with a simple template model.

The detailed analysis reported here was all carried out on data that had been subjected to a 1-36Hz band pass filter prior to processing for other experiments. The blink is a very low frequency component of the EEG and the line segments of the template act, in effect, as low value low pass filters. So the addition of higher frequency EEG should not alter the success of the TEBR method. We later reran the algorithm on new unfiltered data sets and demonstrated similar success to filtered data.

We compared our TEBR method with three different variants of ICA, which is especially good at removing ocular artefacts (Jung *et al.*, 2000; Jung, Makeig, Westerfield, Townsend, Courchesne, and Sejnowski, 2001). TEBR was only slightly better than ICA at more frontal midline sites (1%) but significantly better (>20%) at lateral sites with an average advantage over all of 10%. Importantly, ICA must be used with multiple channels, whereas the TEBR method can be used with just a single channel EEG signal. Further, with standard ICA, you need an expert to spend time viewing the components decomposed by ICA deciding which to remove as the artefact components. Given that TEBR is fully automated, we compared it with ICA, followed with ADJUST to recognise the artefact components automatically. In the version we were using, experimenter time was still needed to remove the components recognised by ADJUST. This could be eliminated to create a fully automated version. However, as noted in section 3.2.2, the results produced across different runs of ICA+ADJUST are not stable. So, we believe that in a range of respects the TEBR method is better than ICA.

One obvious use of TEBR would be in online processing for procedures such as neurofeedback; where only 1 or 2 electrodes may be being used and where calculation time is important. We determined the time costs for each operation in our current method. The conventional desktop processor we were using was an Intel® Core™ i5-6500 CPU @ 3.20GHz with 8.00 GB RAM running a 64-bit Windows 10 Enterprise system. The time to remove an eye blink was 647 ms on average. This means that the procedure could be used for online eye blink removal, where the major time delay

(~1s) would be in the recording of the early components of the blink prior to starting processing. There will be a number of adjustments to our current algorithm (e.g. we have an earlier version of the algorithm that uses slope/rate detection on the initial phase of the blink rather a simple amplitude threshold to detect the peak and does not need individual threshold adjustment) that could optimise it for online processing. In contrast, Infomax ICA procedures based on simultaneous analysis of 16 channels, needed close to 110 seconds to decompose a signal with a duration (and so delay before eye blink correction) of 480s.

4.5 Conclusion

In conclusion, our TEBR method is good at removing eye blink components to leave residual EEG. In terms of data distortion, TEBR appears superior to ICA in our tests. TEBR retains completely unchanged values in areas where there is no eye blink, while ICA methods altered some original brain wave information in non-eye blink areas. In terms of the number of channels, TEBR can be easily used with both single- and multi- channel EEG, while ICA can only be used with multi-channel EEG. In terms of online applications, TEBR should allow online processing for experiments such as neurofeedback training.

However, on occasion, TEBR fails to fit complex multiple blinks (e.g. double-blink, triple-blink). We have already included some adjustments to reduce the number of failures but further improvements should be possible with additional experimenting. Currently, TEBR would also not be expected to deal with the blink compounded with other artefacts (e.g. muscle artefact).

Chapter 5

Feature Extraction

Because the EEG signal is highly complex, unstable and chaotic, to extract and collect features from raw EEG is an important part of EEG analysis. From the cortex to the scalp, the obtained EEG signal is weak, in this situation, signal processing algorithms are used to remove the background noise to increase SNR. They are also used to extract low dimension information from high dimension EEG signals. In this chapter, the feature extraction methods we used in this research are introduced.

5.1 Basic Steps of Feature Extraction

The core of feature extraction is signal processing. Except for the signal processing, there are some other steps used to help extract the features from the original EEG. Figure 5.1 illustrates the basic steps of feature extraction in depressivity prediction models. They are the splitting into epochs, signal processing and feature vector construction.

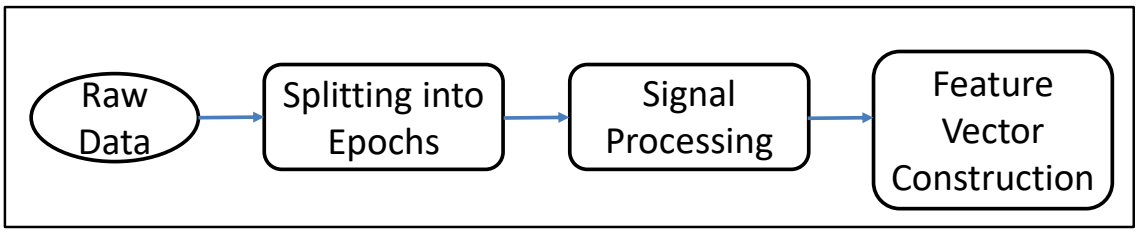

Figure 5.1: The basic steps of feature extraction: splitting into epochs, signal processing and feature vectors construction.

5.2 Splitting into Epochs

In the procedure of data preprocessing (Section 3.4), each data sample in our dataset had already been split into 1-second-length epochs and the first and last epochs of each minute have been discarded (Figure 5.2 upper one). The 1-second length epochs were used in most of our models. On this basis, we started at 128 data points before the original epoch and finished at 128 sample points after the original epoch and extended each epoch to 512 data points (2 seconds) (Figure 5.2 lower one). It means that for each epoch, its first 128 data points overlap with the previous epoch's last 128 data points and its last 128 data points overlap with the following epoch. The scheme that extends the epoch's length into 512 data points was adopted in the models which used FFT as the signal processing algorithm. When implementing the FFT, a Hanning window was applied on the original epochs to get a periodic signal to reduce spectral leakage. After applying the Hanning window (which is a cosine wave), the amplitude smoothly reduces and reaches zero at the edges. Therefore, we overlapped the epochs to offset the loss of data (Figure 5.2), which is the normal procedure with this type of window.

5.3 Signal Processing & Feature Vector Construction

We used two time-frequency domain analysis methods: Fourier transform (FT), continuous wavelet transform (CWT); and one spatial processing algorithm, namely common spatial pattern (CSP) (Blankertz, Tomioka, Lemm, Kawanabe, and Muller, 2008), as our signal processing algorithms. Different models used different signal processing algorithms. For different signal processing algorithms, different methods were designed to construct feature vectors. These algorithms and feature vector construction (FVC) methods are introduced in the following sections.

5.3.1 Fourier Transform

We supposed that depressivity is a mental state which cannot be altered in a short time. Therefore, discrete Fourier transform (DFT) was chosen to calculate the EEG power of each epoch without retaining the time information. DFT is a classic method

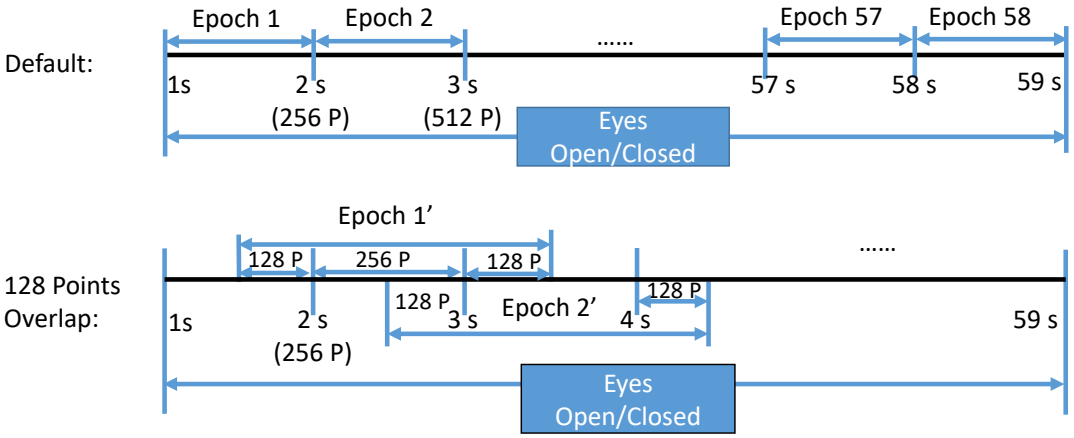

Figure 5.2: Illustration of the epochs splitting. The upper diagram shows the epoch which length is 256 data points. The lower diagram shows the epoch which length is 512 data points, and each epoch overlaps 128 data points with the previous and following epochs. We ranked epochs of eyes open data and eyes closed data separately. Note, our data samples had an 8-minutes signal, but the figure only drew 1-minute signal. The following 7 minutes were omitted because we used the same rule to split epochs as the first 1-minute signal.

to transfer the signal from the time domain into the frequency domain. It is defined as

$$X(m) = \sum_{n=0}^{N-1} x(n) e^{-i2\pi nm/N} \tag{5.1}$$

where $x(n)$ is a discrete sequence of time-domain, $X(m)$ is the discrete frequency-domain sequence. N is the length of $x(n)$.

However, DFT is inefficient, so in our research, Fast Fourier transform (FFT) was used to implement the discrete Fourier transform.

5.3.1.1 Hanning window

EEG signal is a natural signal which does not always include a whole period. If FFT is applied to the raw EEG data directly, spectral leakage will happen. The FFT transform assumes that the measured signal is a finite data set, a continuous spectrum that is one period of a periodic signal. However, the EEG signal is a natural signal and so is a truncated waveform from the original continuous time signal. When the number of periods is not an integer, the endpoints of the time waveform are discontinuous.

These discontinuities cause high-frequency components in the frequency domain of the spectrum that do not exist in the original signal. Therefore, the spectrum obtained by FFT is not the actual spectrum of the original signal. This phenomenon is called spectral leakage (National Instruments, 2019).

In this situation, the EEG signal needs to be multiplied by a finite-length window to reduce leakage. The amplitude of the window varies smoothly and gradually toward zero at the edges. This results in a continuous waveform which meets the FFT's assumptions. The frequency component of a window is a continuous spectrum with a main lobe and several side lobes. The main lobe is centred at each frequency component of the time-domain signal, and the side lobes approach zero (National Instruments, 2019). There are several kinds of windows. In this work, the Hanning window was adopted to reduce the spectral leakage due to the side lobe of the Hanning window being small. The Hanning window is suitable for 95% of cases (National Instruments, 2019). The Hanning window is defined as:

$$\omega(n) = \frac{1}{2}(1 - cos(\frac{2\pi n}{N})) \tag{5.2}$$

N is the length of the original signal, n=0,1,2...N. We applied Hanning window on the original EEG epoch before FFT (See Figure 5.3 (a)).

5.3.1.2 Frequency

FFT transfers a time-domain signal into the frequency domain. After transformation, the length of the result is the same as the original signal's length. The frequency can be computed as:

$$F_n = (n-1) \times \frac{F_s}{N} \tag{5.3}$$

where F_n is the nth point's frequency. F_s is the sampling rate of the original signal. N is the length of the signal. Therefore, the frequency resolution is F_s/N. The FFT returns a two-sided spectrum. Because these two sides are symmetrical, we only retain the first half values. F_1 is the DC component which is outside the scope of our research; therefore, it is also ignored.

5.3.1.3 Amplitude

The FFT returns a spectrum in the complex form $(a + bi)$. The amplitude A_n of each frequency bin F_n can be calculated by Equation 5.4:

$$A_n = \frac{2\sqrt{a^2 + b^2}}{N} \tag{5.4}$$

Because a Hanning Window was applied to the original EEG signal, the estimated power of the original EEG signal was reduced, this was corrected by multiplying the amplitude by a correction factor afterwards. Therefore, A_n was corrected to obtain actual amplitude (M_n) by multiplying the appropriate scaling factor (inverse of the average weight of the Hanning coefficients):

$$M_n = A_n \times (1/mean(Hanning(N))) \tag{5.5}$$

5.3.1.4 Log of power

Then, the amplitude values at each point of frequency domain were squared to obtain the power value. To reduce the skewness resulting from the power transform, each power value is *log* transformed and used to compute a feature.

5.3.1.5 The Example of using Fourier Transform

Figure 5.3 illustrates one example of using the Fourier Transform to process a signal. Our signal processing is based on epochs. Each epoch contains 18 channels. Figure (a) shows a randomly picked epoch/channel (dotted pink line). The segment between two solid red lines lasts 1s (256 sample points), and it is extended to 2s (512 sample points), marked by two dotted green lines. Then a 2s Hanning window was applied on the whole segment. The solid blue line between the two dotted green lines shows the signal after applying a Hanning window; Figure (b) shows the segment (512 sample points) transferred by FFT. The time domain signal was converted to the frequency domain; Figure (c) shows the power of the FFT value; Figure (d) shows the *log* of power.

5.3.2 Feature Vector for FFT

After FFT, each epoch was transferred into frequency domain to generate an $n \times m$ matrix. n is the number of channels, m is the number of frequency slots. In example (5.3.1.5), the frequency resolution is 0.5 Hz, the frequency range is 0.5 to 128 Hz, and the channel number is 18. Then the 18×256 matrix was first reshaped into a vector (length is 4608). Generally, EEG signals are divided into five different frequency bands, their frequency range is from 1 Hz to 80 Hz (Section 2.3.2.2). Therefore, the segment from 1 to 80 Hz of the vector was selected to be the feature vector with a length of 2862 (18×159). Figure 5.4 shows the original matrix and the converted vector.

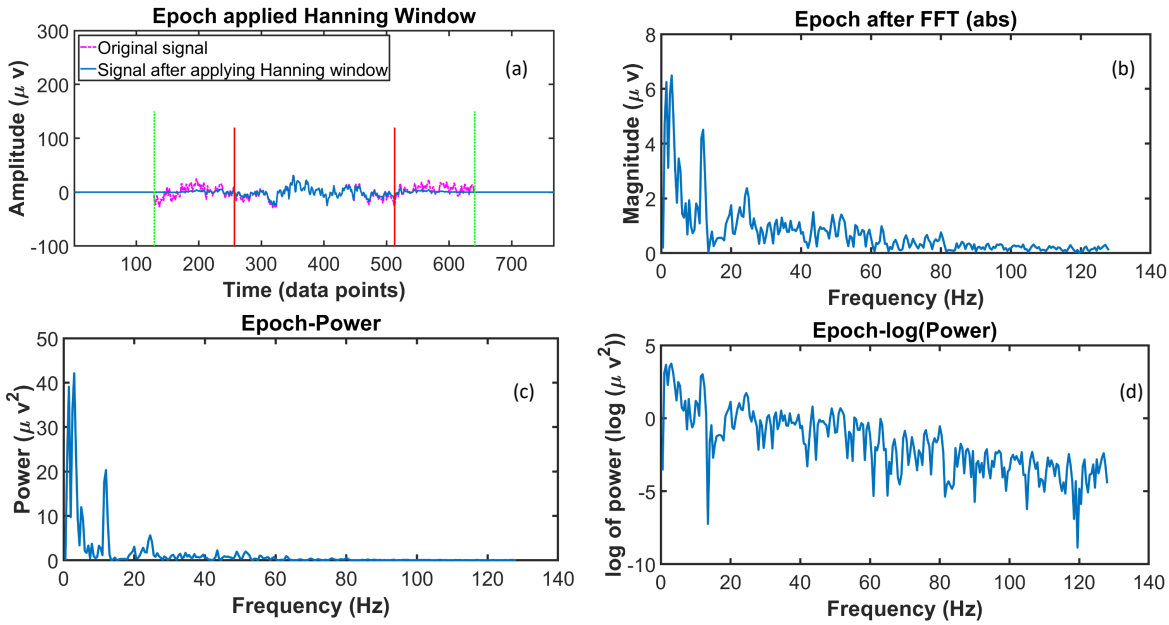

Figure 5.3: Figure (a) shows the epoch picked applied a Hanning window. The epoch between two red solid lines lasts 1s (256 sp), and it is extended to 2s (512 sp), marked by two green dotted lines; Figure (b) shows the segment (512 sp) transferred by FFT; Figure (c) shows the calculated power of FFT; Figure (d) shows the *log* of power.

5.3.3 Wavelet Analysis

Fourier analysis can only derive the frequency domain's information, and the information in time is lost. However, Fourier transform is a good starting point, and the result of FT can be the benchmark for later research. To further analyse the signal, we tried wavelet analysis which cannot only obtain the signal's frequency information, but can also retain time domain information.

There are two kinds of wavelet analysis: continuous wavelet transform (CWT) and discrete wavelet transform (DWT). DWT is suitable to solve problems such as signal compression, de-noising, or signal transmission. However, our primary goal was a time-frequency analysis, so in this research, CWT was used as the signal processing algorithm due to its more fine grained resolution.

CWT calculates the inner products of the signal which need to be analysed and the analysing function to get the similarity between them. Different than FT, the analysing function in CWT can be shifted and compressed or stretched, and it is called wavelet. At different scales (compressed or stretched wavelet) and position (shifted wavelet), the signal and the wavelet are compared, and then a function of two variables is obtained.

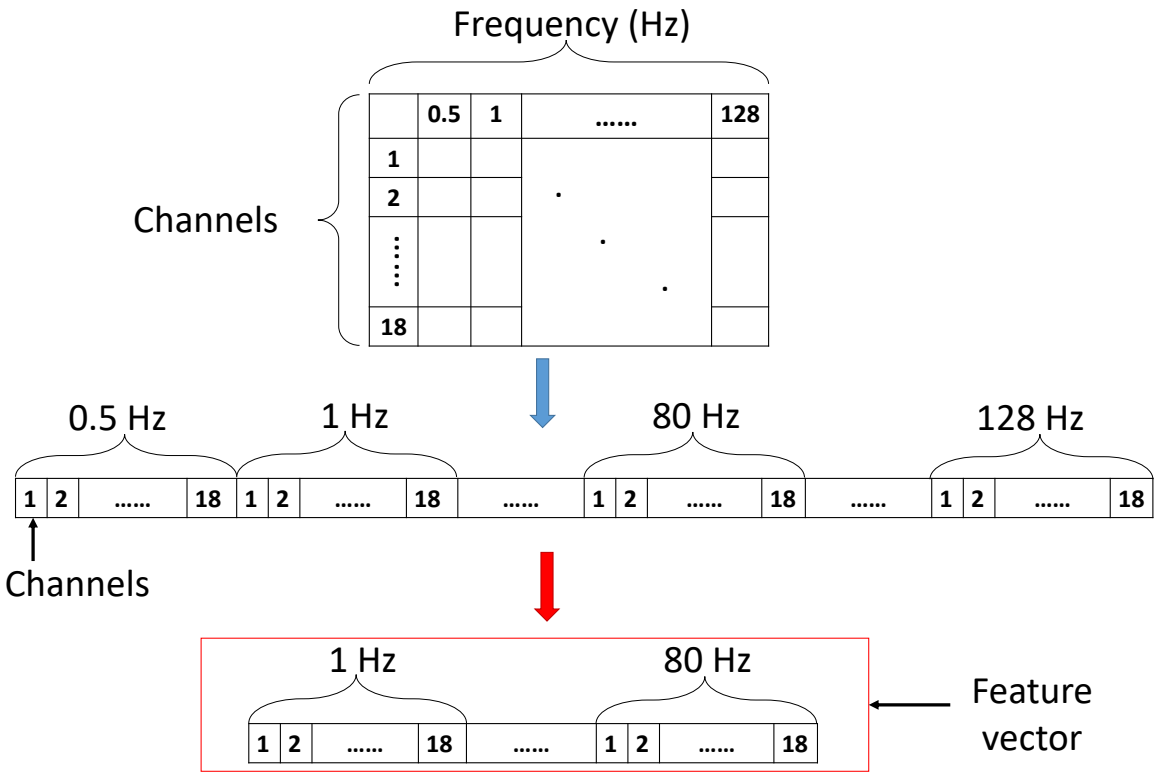

Figure 5.4: This figure shows the feature structure of the FFT-based method. The original structure is 18×256 matrix. 18 is the number of channels, and 256 is the number of frequency slots. Then, the matrix was converted into a vector with a length of 4608 (18×256). Finally, segment from 1 to 80 Hz of the vector was selected to be the feature vector with a length of 2862 (18×159).

The CWT is denoted as:

$$C(s,p;f(t),\psi(t)) = \int_{-\infty}^{\infty} f(t)\frac{1}{s}\psi^*(\frac{t-p}{s})dt \tag{5.6}$$

where s is the scale parameter, so $s > 0$. p is the position parameter, $\psi()$ and $*$ denote the wavelet function and complex conjugate.

5.3.3.1 Scale

The scale concept is an important strength of a wavelet. It helps to analyse the signal in the proper or desired scale.

In general, for signals, scale and frequency have an inverse relationship. In wavelets, if the scale parameter s gets smaller, the wavelet gets more compressed. On the contrary, if the scale parameter s gets longer, the wavelet gets more stretched. In this

condition, the longer portion of the signal is being compared, so the wavelet coefficients measure coarser signal features. The relationship between scale and frequency is shown in Figure 5.5. However, this relationship is general, not precise. Therefore, wavelet transformation will not generate an exact frequency in hertz, but it will generate a pseudo-frequency corresponding to a scale.

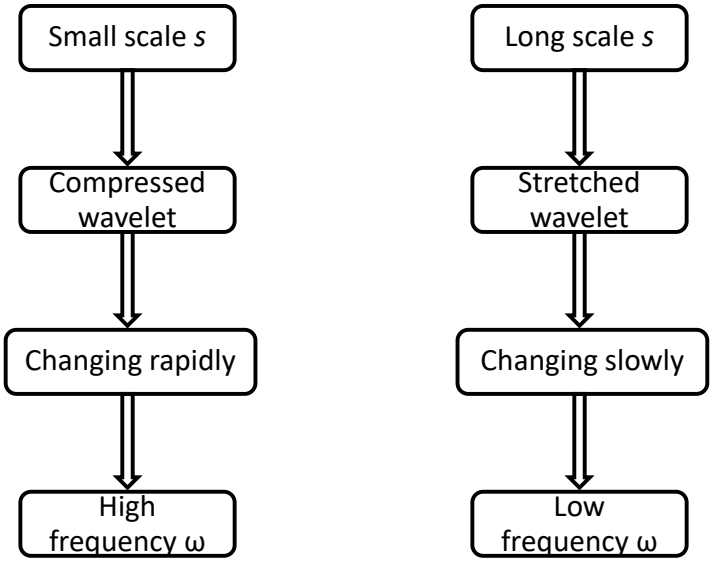

Figure 5.5: This figure illustrates the inverse relationship between scales and frequency.

For example, the sample rate of our data is 256 Hz (1s), and the total scale is set to 1:1:128. The frequency, divided into 128 scales, corresponds to 1-128 Hz. Therefore, the frequency slot is 1Hz.

5.3.3.2 Position

Except for scale, position is another feature of the wavelet. The wavelet is compared with the segment of the signal, and the segment is the same length as the wavelet. Then, the wavelet is shifted to the next position to compare with another segment signal. That is the reason a wavelet can retain time domain information. Figure 5.6 shows a wavelet (left) shift to another position (right).

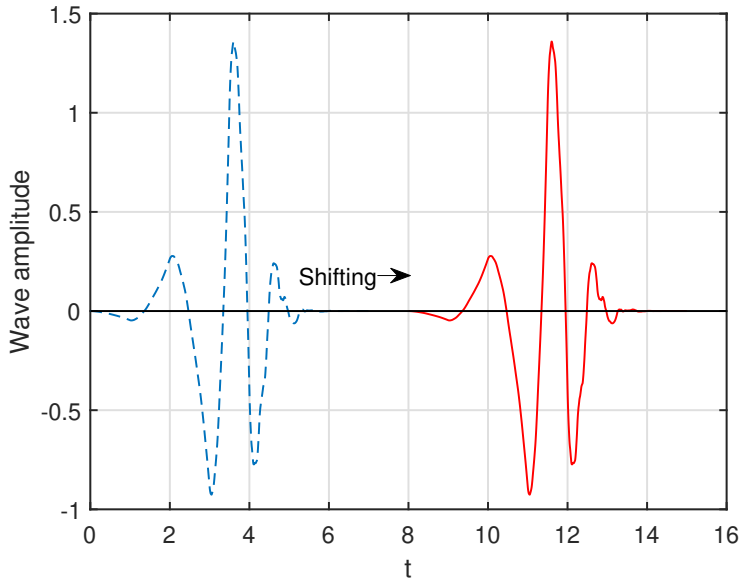

Figure 5.6: This figure Illustrates the wavelet shift to another place.

5.3.3.3 Different wavelets

There are different kinds of wavelets, and they have different shapes and characters. The choice of wavelet will influence the values of the coefficients. In our research, db4 wavelet was adopted because db is an orthogonal wavelet so that it can preserve energy in the analysis stage. Furthermore, this wavelet with smaller support can find closely spaced features.

5.3.3.4 The Example of using CWT

Figure 5.7 shows one example transferred by a continuous wavelet transformation. The upper one shows the original signal. The below one shows the squared value of the wavelet coefficient with time vs approximate frequency.

5.3.4 Feature vector for CWT

After CWT, we had a three-dimensional (3D) array for each epoch. The first dimension was 'Time', the second dimension was 'Frequency (Approximately)', and the third was 'Channel'. To convert the entire matrix into a 1-D feature vector can result in a vector which contains a vast amount of data. To decrease the dimensionality of the feature vectors, we used two different methods to create feature vectors.

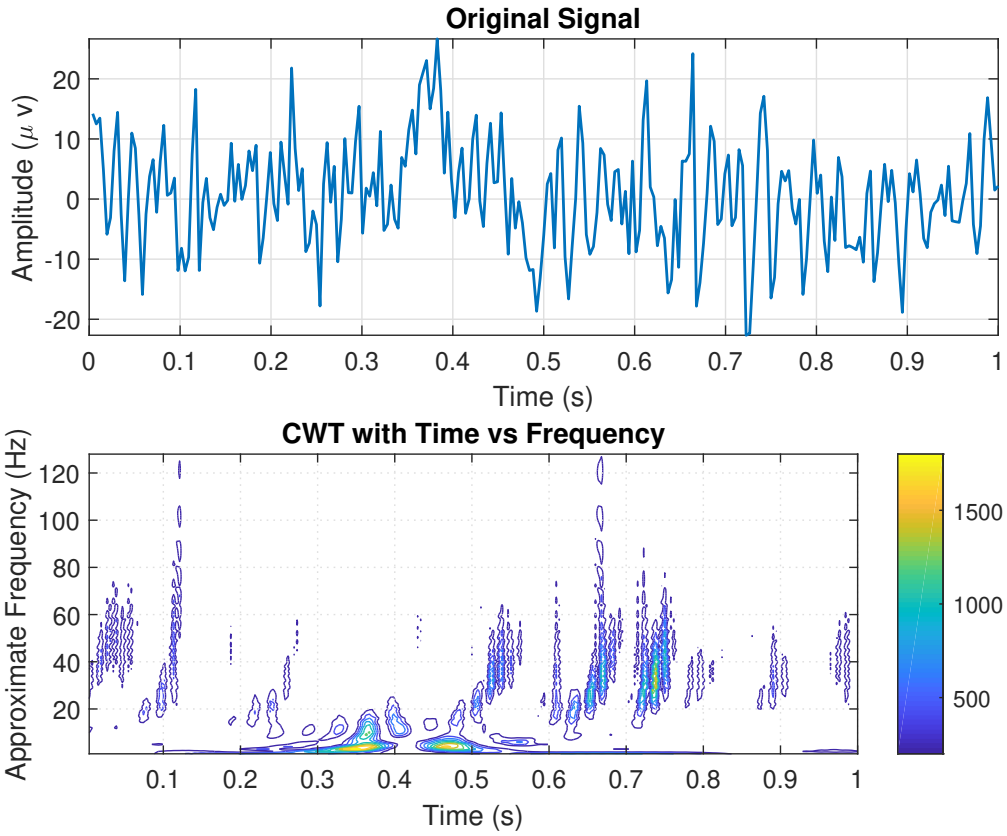

Figure 5.7: Top: the original signal. Bottom: the squared value of the wavelet coefficients with time vs approximate frequency.

1) Accumulated energy All values along the time domain were averaged to calculate accumulation of energy in each frequency slot:

$$E_j = \frac{1}{T}\sum_{i=1}^{T} d_j^2(i) \tag{5.7}$$

where, E_j is the accumulated energy of jth frequency. T means there are T time points. $d_j^2(i)$ $(i = 1, ..., T)$ are the squared wavelet coefficient values in the jth frequency.

This method is similar to the FT because the 'Time' dimension was removed. Figure 5.8 as an example shows that each frequency slot (frequency resolution is 1 Hz) has one value. As the same as FFT, only the segment from 1 to 80 Hz was selected. The 18×80 matrix was reshaped into the CWT's feature vector (length is 1440). It used the same feature vector construction method displayed in Figure 5.4.

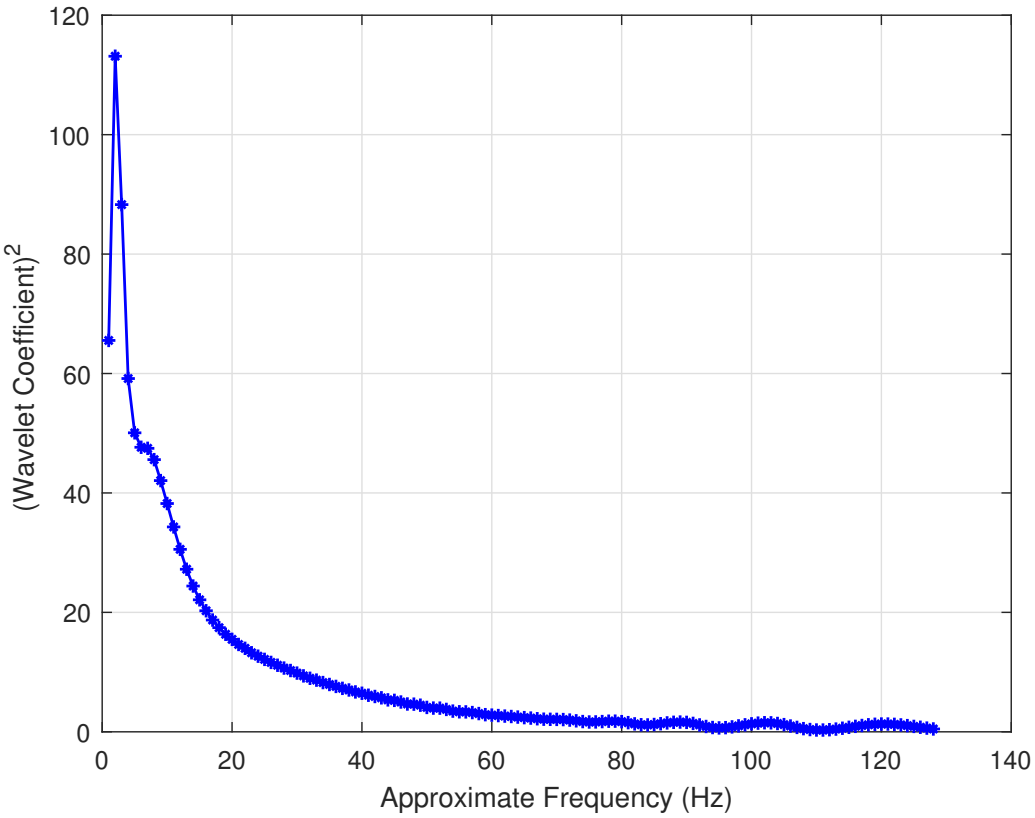

Figure 5.8: All values along the time domain were averaged to obtain one value in each frequency slot. This method actually is similar to the FT because the 'Time' dimension was removed.

2) Top five sudden changes The second figure of Figure 5.7 shows that each time/frequency intersection point has one squared wavelet coefficient value. Large wavelet coefficient values reflect sudden changes in the EEG signal. This kind of feature (sudden change) has been used to detect epileptic seizures (Faust, Acharya, Adeli, and Adeli, 2015; Ahammad, Fathima, and Joseph, 2014; Fathima, Bedeeuzzaman, Farooq, and Khan, 2011). Here, we used the sudden change as a feature to see if it relates to depressivity. For epileptic seizure recognition, usually only the maximum value in each frequency slot is used to establish the feature vector (Jahankhani, Kodogiannis, and Revett, 2006; Ahammad *et al.*, 2014; Fathima *et al.*, 2011). To retain more information, we extended the selected wavelet coefficients from just one to the top five which we deem to be a reasonable, if arbitrary, number. Therefore, the top five maximum squared

wavelet coefficient values in the time domain in each frequency slot (frequency range: 1-80 Hz) were picked to form a feature vector (Figure 5.9). For an epoch, the length of feature vector is $5 \times 18(channel) \times 80(Hz) = 7200$.

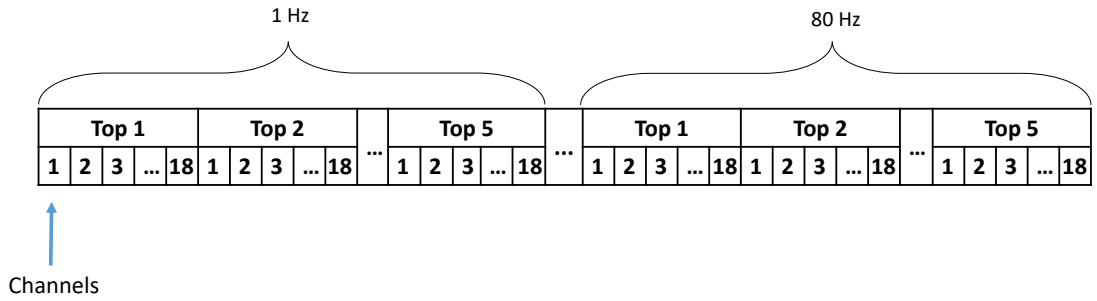

Figure 5.9: This figure shows the feature vectors of the CWT-based method: top five maximum values. In this feature vector, the top five maximum values in the time domain in each frequency slot (frequency range: 1-80 Hz) were picked to form a feature vector.

5.3.5 Common Spatial Pattern (CSP)

Other than two time-frequency signal processing algorithms, one spatial processing algorithm (common spatial pattern) was also adopted in this research. The method of CSP was introduced in the field of EEG analysis by Koles, Lazar, and Zhou (1990) who used the method to classify normal and abnormal EEGs (Koles, Lind, and Flor-Henry, 1994). Müller-Gerking, Pfurtscheller, and Flyvbjerg (1999) used CSP to classify EEG in movement task and localise the most related EEG's sources. The algorithm of CSP is based on two variance matrices simultaneously diagonalised (Wang, Gao, and Gao, 2006), and can transform high-dimensional EEG data into low-dimensional data, which are more distinguishable between two distinct sets of data that are obtained under different conditions (e.g. left-hand movement versus right-hand movement in motor-imagination EEG). It leads to maximum differences in variance between the two sets of variance matrices. Thus, it is very suitable for extracting the contrasting features of data. In this research, we used CSP to classify high and low depressivity EEG epochs to evaluate the depressivity of participants and localise the sources of EEG.

5.3.5.1 Algorithm

The algorithm of CSP (Müller-Gerking *et al.*, 1999) is presented in this section.

1. First, the normalised covariance matrix is calculated for each selected epoch:

$$C_h^i = \frac{E_h^i (E_h^i)'}{\text{trace}(E_h^i (E_h^i)')} \tag{5.8}$$

   where matrix E_h^i is the raw data of epoch i, condition h (High depressivity). The structure of the matrix is $N \times T$, N is the number of channels (18 in our case), and T is the number of data points in each epoch. E' is the transposed matrix of E, and trace(M) is the sum of the diagonal elements of M. Also, C_l^i denotes the corresponding normalised covariance matrix for the trial i of condition l (Low depressivity).

2. Second, the averaged normalised covariance matrices are calculated by the average over all epochs in each condition,

$$C = \overline{C^i}, \tag{5.9}$$

   C_h is the averaged normalised covariance of High depressivity EEG. Likewise, the averaged normalised covariance of Low depressivity EEG, C_l is calculated using the selected low score EEG epochs.

3. Third, the composite covariance matrix is factored as:

$$C_s = C_h + C_l = B_s \gamma_s B_s' \tag{5.10}$$

   where B_s is a matrix of eigenvectors and γ_s is a corresponding diagonal matrix of eigenvalues. The eigenvalues here are assumed to be sorted in descending order.

4. Fourth, we then transform individual $C_h$and C_l as:

$$S_h = W C_h W' \tag{5.11}$$

$$S_l = W C_l W' \tag{5.12}$$

   where $W = \sqrt{\gamma_s{}^{-1}} B_s'$ is the whitening transformation, which equalises the variances in the space spanned by the eigenvectors in B_s.

5. Fifth, since $S_h + S_l = WC_sW' = I$ (I is the identity matrix), S_h and S_l share the same eigenvectors, there is:

$$S_h = U\gamma_h U' \tag{5.13}$$

$$S_l = U\gamma_h U' \tag{5.14}$$

and $\gamma_h + \gamma_l = I$. Since $\gamma_h + \gamma_l = I$, the eigenvectors with the largest eigenvalues for S_h have the smallest eigenvalues for S_l and vice versa. The transformation of whitened EEG epochs onto the eigenvectors corresponding to the largest eigenvalues in γ_h and the smallest eigenvalues for γ_l. If the difference between γ_h and γ_l is very big, it is better to separate the variances of high score and low score EEG data.

6. Finally, the projection matrix $P' = U'W$ (spatial filter) is applied in the following equation to get the feature matrix Z^i for each epoch E^i.

$$Z^i = P'E^i \tag{5.15}$$

From Equation 5.15, we can see that the main idea of the CSP method is to use a linear transform to project the multi-channel EEG data into another space with a projection matrix (Ge, Wang, and Yu, 2014).

Inverting Equation 5.15,

$$E^i = (P')^{-1}Z^i \tag{5.16}$$

It shows that the raw EEG E^i is decomposed into the CSP patterns (columns of $(P')^{-1}$) and the expansion coefficients Z^i. The patterns $(P')^{-1}$ can be seen as a form of source distribution. Z^i is the corresponding source wave form matrix and can be used as a feature matrix. The first row of Z^i corresponds to the first direction in the new space which contains the most related information of the condition of High depressivity, and the last row of Z^i corresponds to the first direction in the new space which contains the most related information of the condition of Low depressivity. The second and the second-to-last directions are the second best directions for classifying High and Low depressivity, and so on for the other directions in the list.

The patterns as source distribution can be directly treated as biomarkers extracted from EEG. Figure 7.10 shows an example of a CSP pattern. It shows the locations that are related to High depressivity.

Because CSP is used to distinguish two sets of data, it will be only used in classification-based models.

5.3.5.2 Implementation

In the training phase, the CSP spatial filter (P' in Equation 5.15) was generated. This filter was then applied to extract features (Z^i in Equation 5.15) in both the training and testing data.

Training phase In the training phase, training data — selected High depressivity data samples and selected Low depressivity data samples — were pooled into a high group and low group of epochs, respectively. These groups were then submitted to the CSP algorithm to get the CSP spatial filter. Then, each epoch of training data was filtered with CSP spatial filter to get the feature matrix to train the classification model (Figure 5.10). The first and last rows of the matrix contain the information most relevant to the two distinct conditions. Then, the second and the second to last rows have less relevant information. The middle rows have the least information. In this thesis, the first n and the last n rows of the feature matrix were saved as features. n varied from 1 to 5 and was set empirically. Our results reported in Section 7.5.3.1 show that after $n = 3$, the R^2-values increased slowly, suggesting that $n = 5$ is enough to include the bulk of valid information.

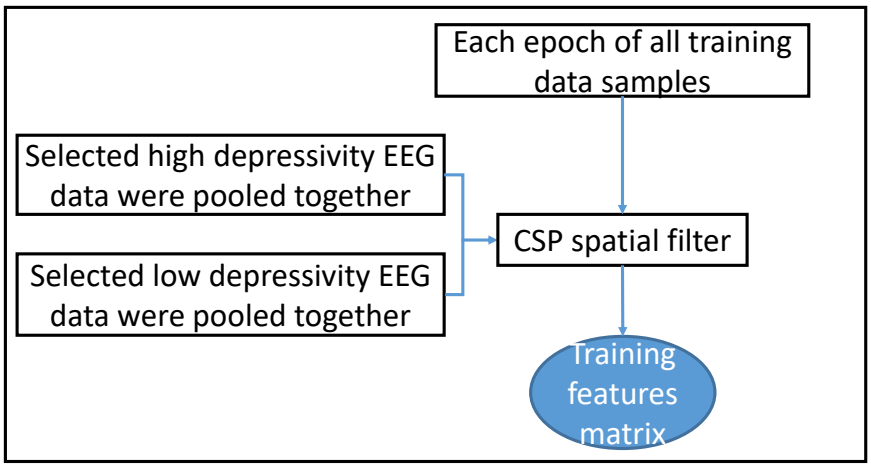

Figure 5.10: This figure shows the steps of the signal processing of CSP in the training phase. All were pooled into a high group and low group of epochs, respectively. These groups were then submitted to the CSP (common spatial pattern) algorithm to get the CSP spatial filter. Then, each epoch of training data was filtered with a CSP spatial filter to get the features to train the classification model.

Testing Phase In the testing phase, the step of calculating the CSP spatial filter was omitted, and we used the learnt filter calculated at the training step.

The scheme of CSP in the testing phase is shown in Figure 5.11. For this, we obtained features of each epoch of each data sample separately.

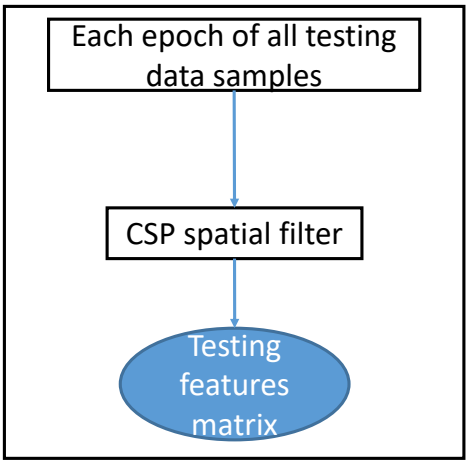

Figure 5.11: CSP in the testing phase. In this phase, we obtained the feature of each epoch of an individual's data without pooling all the data together.

5.3.6 Feature Vector for CSP

As described in Section 5.3.5.1, the shape of the feature matrix of CSP for each epoch is $N \times T$, the same size as raw data. The first and last n rows of a feature matrix contain the most relevant information for each class, so these $2 \times n$ rows were picked up. Then, we used two methods to construct feature vectors.

1) Whole feature rows The first method is to pick up the first n and last n rows of the feature matrix to put them together to align a feature vector. n varies from 1 to 5.

Figure 5.12 shows the features extracted from one experiment. The experiment used the data described in Section 3.2 of Chapter 3. The size of the feature matrix is $N \times T$. Here, we only used two rows (the first row and last row of the feature matrix). The length of T is 256. Therefore, the length of each epoch's feature vector is $2 \times 256 = 512$. The feature vector marked by red circles of Figure 5.12

shows the averaged feature vectors of all high depressivity training data. The feature vector marked by blue crosses of Figure 5.12 shows the averaged feature vectors of all low depressivity training data. It shows that High depressivity has a relatively high variance in the first half of the feature vector, while low depressivity has a relatively high variance in the second half of the feature vector.

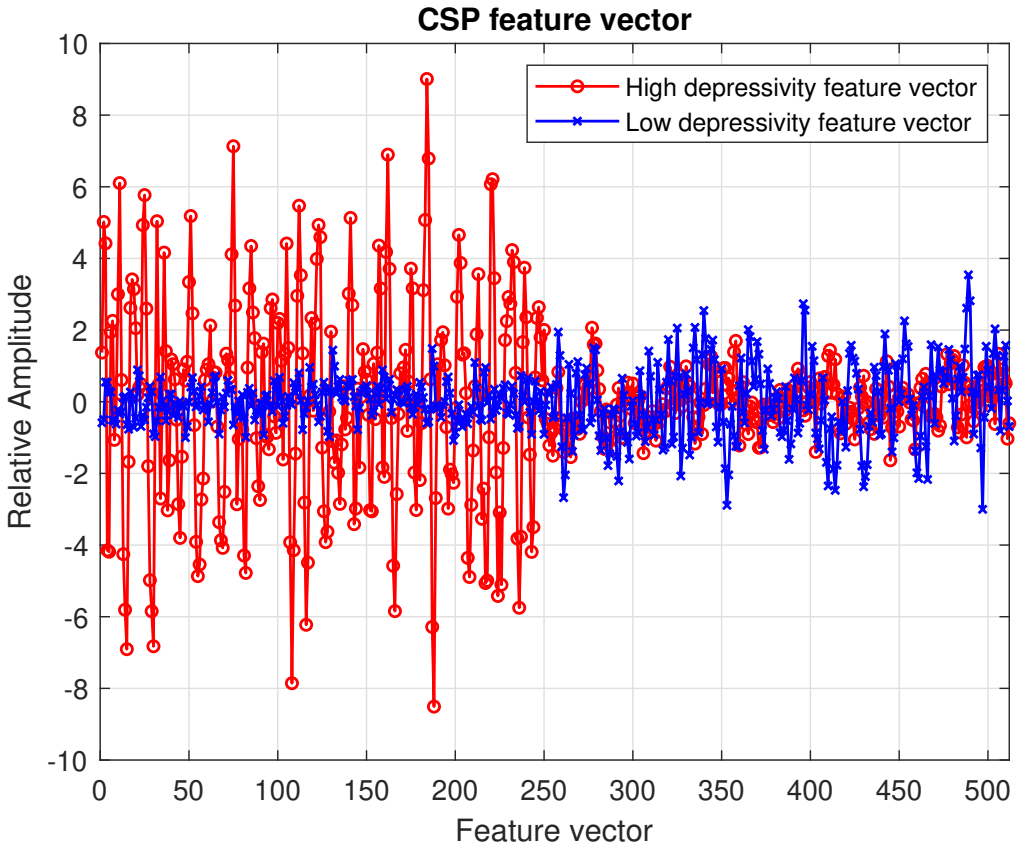

Figure 5.12: This figure shows the comparison of averaged feature vectors of high depressivity data and low depressivity data. The solid red line shows the averaged feature vectors of all high depressivity training data. The dotted blue line shows the averaged feature vectors of all low depressivity training data. It shows that the High depressivity has a relatively high variance in the first half of the feature vector, while low depressivity has a relatively high variance in the second half of the feature vector.

2) Variance of rows The features extracted from the first method show the clear difference of variance. Therefore, the second method is to pick up the first n and last n rows from the feature matrix and calculate each row's variance. The

variance of the row p is denoted by var_p. The feature vector for an epoch consists of the $2n$ variances var_p, $(p \in [1, n] \cup [N - n + 1, N])$. Each var_p is normalised by all variances and log transformed (Müller-Gerking *et al.*, 1999):

$$f_p = log(\frac{var_p}{\sum_{p=1}^{2n} var_p}) \tag{5.17}$$

Compared to using raw feature rows, the length of a feature vector obtained by this method is shorter because there is only one value for one feature row.

5.4 Conclusion

This chapter introduces the feature extraction flow of the EEG analysis in our study. The first step is the epoch's split, and the raw data is split into equally long epochs. The second step is signal processing. In this research, we used three signal analysis methods: FFT, CWT and CSP. The last step is feature vectors construction. In this step, the feature vectors which are fed into predictive models are constructed. It needs to be noted that in the procedure of designing classification-based models, a new feature extraction method – FFT combines CSP was designed. The details will be introduced in Section 7.6.2.

Chapter 6

Predictive Methods

This chapter introduces the predictive methods used in this research: one neural network (Learning Vector Quantisation, LVQ); three ensemble learning methods (Adaptive Boosting, AdaBoost; Least Squares Boosting, LSBoost; and Random Forest, RF); and one improved least squares regression method (the Least Absolute Shrinkage and Selection Operator, Lasso). These methods were used to establish different depressivity prediction models.

6.1 Introduction

The aim of our research is not to simply distinguish a 0/1 classification (depression or not depression), but what we try to predict is the depression level (depressivity). As mentioned before, two major types of models were designed, and they are based on two ideas: the appearance of some specific EEG patterns, which indicate High depressivity, increases in number with a higher depressivity; and the EEG patterns which vary with different levels of depression.

To realise these two design ideas, several predictive methods were chosen. The classification-based methods are used to realise the first idea. The regression-based methods are used to realise the second idea. They are introduced in the following sections.

In this chapter, $Tr = (x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ is defined as a training dataset. Each x_i belongs to some domain or instance space X, where $x_i = (x_{i1}, .., x_{ip})^T$, and y_i is either a class label or a numerical response. These definitions are used for introducing algorithm flows of predictive methods.

6.2 Neural Network (NN)

Artificial neural network (ANN) is widely used in EEG classification field (See Table 2.2). Although it is called a neural network, it is not a concept in physiology but a mathematical model. It has this name because it simulates the structure of the brain's neural network for information processing. ANN is also known as simply neural network (NN).

NN consists of a large number of artificial neurons, which are connected with each other by the network's topology. Like biological neurons, an artificial neuron consists of an input signal and output signal. An artificial neuron receives an input signal (x), and then the signal will be processed and transferred to the additional artificial neurons connected to it.

6.2.1 Learning Vector Quantisation (LVQ)

There are different kinds of neural networks. In our research, the Learning Vector Quantisation (LVQ) neural network algorithm (Kohonen, 1990, 1995) was selected. The structure of LVQ is simple and LVQ is easy to implement (Guo and Wong, 2013). It is also frequently used for the classification of EEG data (Poulos, Rangoussi, and Alexandris, 1999; Vuckovic, Radivojevic, Chen, and Popovic, 2002; Poulos, Rangoussi, Alexandris, Evangelou, *et al.*, 2002; Ocak, 2008; Dhar, Dhar, Dey, Pal, Saha, and Goswami, 2017; Pradhan, Sadasivan, and Arunodaya, 1996). To compare with previous studies, we used the LVQ method as a baseline, with no specific expectations of its performance.

LVQ is a forward supervised learning network which has three layers (Figure 6.1). They are called the input layer, the competitive layer and output layer, respectively. The competitive layer has several hidden neurons and the output layer has several output neurons. The network is fully connected between the input layer and the competitive layer, and partially connected between the competitive layer and the output layer. When an input epoch is sent to the network, the hidden neuron which is closest to the input epoch wins the competition and produces a 1 and all other hidden neurons produce 0. The output neuron connected to the winning neuron also produces a 1, while the other output neurons produce 0. The output neuron that produces the 1 gives the class of the input pattern (Andina and Pham, 2007).

There are three types of LVQ: LVQ1, LVQ2, and LVQ3. LVQ2 and LVQ3 were proposed to improve LVQ1 empirically (Camastra and Vinciarelli, 2001). Since LVQ1

is the basis of these three options, and they yield almost similar accuracies in most tasks (Kohonen and Maps, 1995), LVQ1 was chosen for our project.

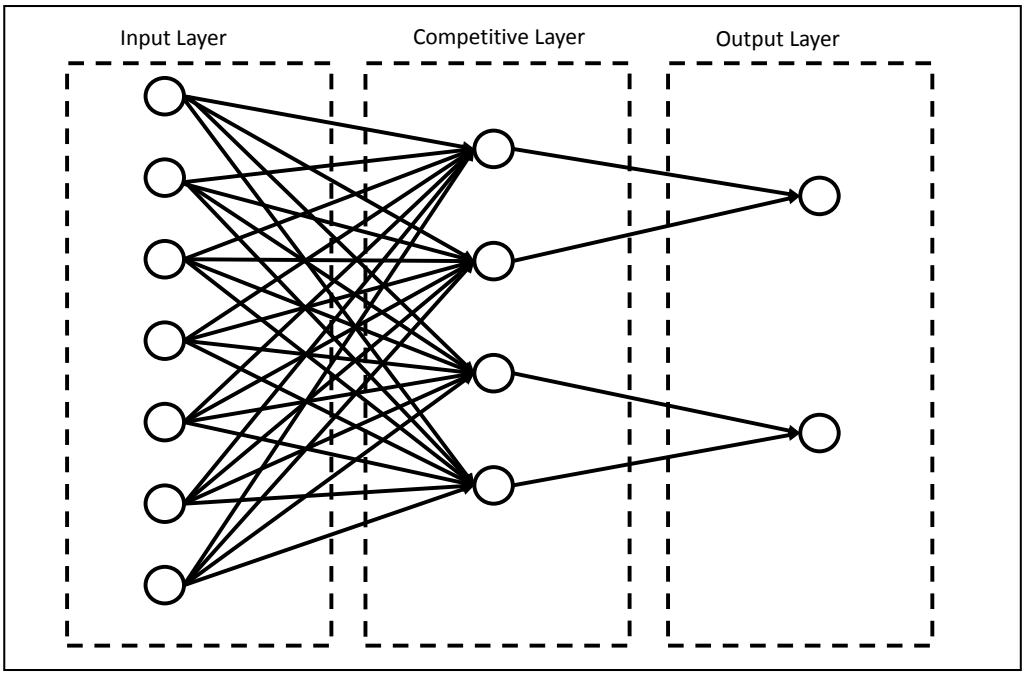

Figure 6.1: LVQ neural network has three layers: Input Layer, Competitive Layer, Output Layer.

Training flow of LVQ1

- **Step One:** Initialise the learning rate η and the weight matrix W between the first and second layers.

- **Step Two:** Present an input vector x_i to the network and calculate the Euclidean distance between the input vector x_i and the hidden neurons in the competition layer:

$$d_{ij} = ||x_i - W_{ij}||, j = 1, 2, \cdots, C \tag{6.1}$$

  where, C is the number of the hidden neurons in the competition layer; W_{ij} is the weight between the x_i and the jth neuron.

- **Step Three:** Find the winning neuron m which corresponds to the minimum distance d_{im}.

- **Step Four:** If the associated target class of the winning neuron is the same as the predetermined class y_i of the input vector x_i, the classification is correct, the weight of the neuron m is adjusted using Equation 6.2; otherwise, the classification is incorrect, and the weight of the neuron m is adjusted using Equation 6.3.

$$W_m(\text{new}) = W_m(\text{old}) + \eta(x_i - W_m(\text{old})) \tag{6.2}$$

$$W_m(\text{new}) = W_m(\text{old}) - \eta(x_i - W_m(\text{old})) \tag{6.3}$$

- **Step Five:** Return to Step Two with a new input vector and repeat Step Two, Three and Four until all input vectors are correctly classified or a stopping criterion is met.

In this project, the learning rate η was set as 0.01. Each row of the initial weight matrix W was set to have the same value (the median of the input ranges). We set two stopping criteria: the error tolerance (0.1) and the maximum number of iterations (100). The input vectors were the feature vectors extracted from EEG for training; the two target output classes were the High depressivity EEG class and Low depressivity EEG class.

6.3 Ensemble Learning

Apart from NN, we also tried ensemble learning which is a powerful tool for data analysis. Ensemble learning is becoming increasingly popular and has been applied to a variety of domains, such as malware detection (Zhang, Yin, Hao, Zhang, and Wang, 2007; Shabtai, Moskovitch, Elovici, and Glezer, 2009) and intrusion detection (Locasto, Wang, Keromytis, and Stolfo, 2005; Giacinto, Perdisci, Del Rio, and Roli, 2008). It is also effectively used in face recognition(Mu, Lu, Watta, and Hassoun, 2009), speech recognition(Krajewski, Batliner, and Kessel, 2010) and facial emotion recognition(Ithaya Rani and Muneeswaran, 2016; Rani and Muneeswaran, 2017, 2018).

The basic concept of ensemble learning is combining many weak learners into a strong ensemble learner. Therefore, it can average out biases, reduce variance and it is unlikely to overfit. The most important advantage of ensemble learning is its easy interpretation, and we can easily find the most important features the ensemble picks. These features are potential biomarkers we try to find. That is why we chose ensemble learning as our predictive method.

There are five components of an ensemble learner. They are input data, the corresponding labels, ensemble aggregation method, weak learner and the number of weak learners.

In this research, we had a dataset including the PID5-d scores and EEG data of participants. EEG data are the input data to be analysed, and the PID5-d scores are the corresponding labels. There are different ensemble aggregation methods in ensemble learning. Some are suitable for classification, while others are suitable for regression. They will be discussed separately in the following sections. The weak learner is also an important part of the ensemble learning. We chose decision tree as the weak learner in the research because it is a most popular algorithm, which can be easily implemented and interpreted.

6.3.1 Ensemble Aggregation Methods

In this research, we used three different ensemble aggregation methods:

1. Adaptive Boosting (AdaBoost): Suitable for classification problems;
2. Least Squares Boosting (LSBoost): Suitable for regression problems;
3. Random Forest: Suitable for both regression and classification problems.

6.3.1.1 Adaptive boosting (AdaBoost)

The AdaBoost algorithm is a widely used ensemble aggregation method for classification.

Introduction AdaBoost was proposed by Freund and Schapire (1997). The basic idea of AdaBoost is combining multiple weak learners into a strong learner. The key to realise this idea is to change the weights of samples in each round of iterative training, and the samples misclassified by the previous weak learner will be strengthened. The weighted samples will be used again to train the next weak learner. The procedure continues until the predetermined error rate, or the predetermined maximum number of iterations is reached. Finally, the combination of the weak learners is the strong learner that we expected.

Algorithm flow

- **First Step:** the weight distribution of training data was initialised:

$$D_1 = (w_{11}, w_{12}, ...w_{1i}, ...w_{1n}), i = 1, 2, ...n, \tag{6.4}$$

where $w_{1i} = \frac{1}{n}$, and in the beginning, each training sample is given the same weight.

- **Second Step:** the algorithm starts m $(m = 1, 2, \cdots, M)$ rounds iteration training.

1. Use weighted training dataset to learn and get the weak leaner $G_m(x)$ (in this method, it was called classifier). The strategy of choosing the weak classifier is according to the lowest error rate.
2. Calculate the classification error rate (ε_m) of $G_m(x)$.

$$\varepsilon_m = P(G_m(x_i) \neq y_i) = \sum_{i=1}^{n} w_{mi} I(G_m(x_i) \neq yi) \tag{6.5}$$

where, m is the order of iteration.

The error rate ε_m of $G_m(x)$ in the training dataset is the sum of the weights of the samples misclassified by $G_m(x)$.

3. Calculate the parameter α_m of $G_m(x)$:

$$\alpha_m = \frac{1}{2} log \frac{1 - \varepsilon_m}{\varepsilon_m} \tag{6.6}$$

α_m represents the importance of $G_m(x)$ in the final classifier (strong learner). α_m increases with the decrease of ε_m which means that if the classifier has a lower error rate, it plays a more important role in the final strong classifier.

4. Update the weight distribution of the training dataset for the next iteration:

$$D_{m+1} = (w_{m+1,1}, w_{m+1,2}, ...w_{m+1,i}, ...w_{m+1,n}), i = 1, 2, ...n, \tag{6.7}$$

where,

$$\begin{aligned} w_{m+1,i} &= \frac{w_{mi}}{Z_m} \exp\left(-\alpha_m y_i G_m(x_i)\right), i = 1, 2, ...n \\ Z_m &= \sum_{i=1}^{n} w_{mi} \exp\left(-\alpha_m y_i G_m(x_i)\right) \end{aligned} \tag{6.8}$$

Z_m is a normalisation factor to make the $G_m(x_i)$ to be a probability distribution. The aim of weight updating is to increase the weight of misclassified samples by $G_m(x_i)$ and decrease the weight of correctly classified samples. Therefore, the next weak learner will focus on 'hard' samples.

- **Third Step:** Assemble weak learners (classifiers) to obtain the final strong learner $G(x)$:

$$\begin{aligned} f(x) &= \sum_{m=1}^{M} \alpha_m G_m(x), \\ G(x) &= \text{sign}(f(x)) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right) \end{aligned} \tag{6.9}$$

where $f(x)$ is the combination of weighted weak learners. $G(x)$ is a weighted majority vote of weak learners' combination. From a statistical perspective, AdaBoost can be viewed as an additive model with exponential loss function in a forward stagewise manner (Friedman, Hastie, Tibshirani, *et al.*, 2000).

6.3.1.2 Least Squares Boosting (LSBoost)

The second ensemble aggregation methods used in this research is least squares boosting (LSBoost). It is a regression-based method. Like AdaBoost, the performance of LSBoost is also boosted iteratively by adding weak learners until a strong learner is generated. As mentioned previously, AdaBoost can be viewed as an additive model with exponential loss function in a forward stagewise manner. Similarly, LSBoost can be viewed as an additive model with squared error loss function. In order to minimise the loss function, the prediction of the current round should be as close as possible to the residual of the previous round (Friedman, 2001):

$$(\rho_m, a_m) = \text{argmin}_{a,\rho} \sum_{i=1}^{N} [y_i - F_{m-1}(x_i) - \rho h(x_i; a)]^2 \tag{6.10}$$

where, m is the number of the current round; $h(x; a_m)$ is the trained weak learner in this round; $F_{m-1}(x)$ is the aggregation learner obtained from the last round; and y is the corresponding output value.

Then, the weak learner is added to the previous aggregation learner (Friedman, 2001):

$$F_m(x) = F_{m-1}(x) + \rho_m h(x_i; a_m) \tag{6.11}$$

where, first $F_0(x) = \bar{y}$. The procedure stops when the predetermined error rate is reached, or the predetermined maximum number of iterations is reached.

6.3.1.3 Random Forest (RF)

The third ensemble aggregation method used in this research is random forest (RF). As a popular and powerful algorithm, it is a type of bootstrap aggregation (bagging).

As an ensemble method, bagging is also a technique that combines several weak learners to obtain a strong learner. There are three essential steps in the bagging method (Breiman, 1996):

1. Create k training subsets. Each subset includes n training samples which were randomly extracted from the original training dataset with replacement. k training subsets are independent of each other. Each training subset is used to train a weak learner. The training samples not included in the kth subset are called out-of-bag samples.

2. Train k weak learners. For k training subsets, k weak learners were obtained. All learners are equally important.

3. Bagging can be used in both classification and regression situations. For the classification problem: results are generated by voting; for the regression problem, the mean value of the predicted results of k models is taken as the final prediction result.

Random forest is a type of bagging, and the weak learners of RF are decision trees. There is a risk in the original bagging method that the structures of some decision trees are similar because all input variances are considered to split nodes. This leads to the resulting predictions of some weak learners being highly correlated. RF is different from the original bagging method. In RF, only a random subset of the input variances is selected at each split node in a tree (Breiman, 2001). So the resulting predictions are less correlated. Therefore, the performance of RF is better than the original bagging. In this research, the square root of the number of variables was randomly picked to split nodes for classification and one-third of the number of variables for regression.

6.3.2 Weak Learner (Decision Tree)

In the ensemble learning method, there is an important component: the weak learner. The basic strategy of the ensemble learning method is adding a weak learner in each iteration and eventually obtaining a strong learner.

In this research, the decision tree is adopted as weak learner for all ensemble aggregation methods. There are four reasons for using the decision tree. Firstly, the decision tree is a commonly used supervised machine learning method in data mining (Rokach and Maimon, 2008). Secondly, the decision tree is easy to understand and interpret, and it can easily estimate the importance of each variable. Thirdly, the decision tree

is relatively fast to construct. Finally, the decision tree is resistant to the inclusion of many irrelevant predictor variables and immune to the effects of predictor outliers (Friedman, Hastie, and Tibshirani, 2001).

There are several kinds of decision trees. In this research, we used CART (Classification and Regression Trees) which can be used in both discrete and continuous conditions. The characteristic of CART is that it constructs binary trees.

6.3.2.1 An example of CART

Figure 6.2 is an example of CART of distinguishing a container from being a Bowl or Cup. There are three elements in the decision tree: decision node, leaf node and branch. Each decision node represents an input variable (e.g. Height). As a binary tree, each node has two branches, and each branch denotes the response value of a test. Each leaf node represents a class label (e.g. Bowl) or numerical value. The topside node is called the root node, and it is the most important node in a tree.

This tree predicts the container's class based on three independent variables: Without handle, Height, and Diameter. Start at the topside node, which is represented by a rectangular box. The first decision is whether the container is without a handle. If so, follow the right branch, and check that the leaf node (ellipse box) shows the container as a bowl. If the container has a handle, then follow the left branch to the lower rectangular node. Here, the tree asks if the height of the container is smaller than 6cm. If so, then follow the right branch to check that the leaf node (ellipse box) shows the container as a bowl. If not, then follow the left branch again to the lower rectangular node. Here, the tree asks if the diameter of the container is bigger than 5cm. If so, then follow the right branch to check that the leaf node (ellipse box) shows the container as a bowl. If not, then follow the left branch to check that the leaf node shows the container as a cup. Given the values of the independent variables (Without handle, Height, and Diameter) of a container, we know the container's class (Bowl or Cup) as long as we follow the rules indicated by the tree (Figure 6.2).

6.3.2.2 Create standard CART

This section introduces the steps to create a CART like the tree shown in Figure 6.2. We first traverse all split points of all variables to find the best split strategy to split all samples into two parts at the root node. Then again, we traverse all split points of all variables to find the best split strategy to split one or both of these parts into

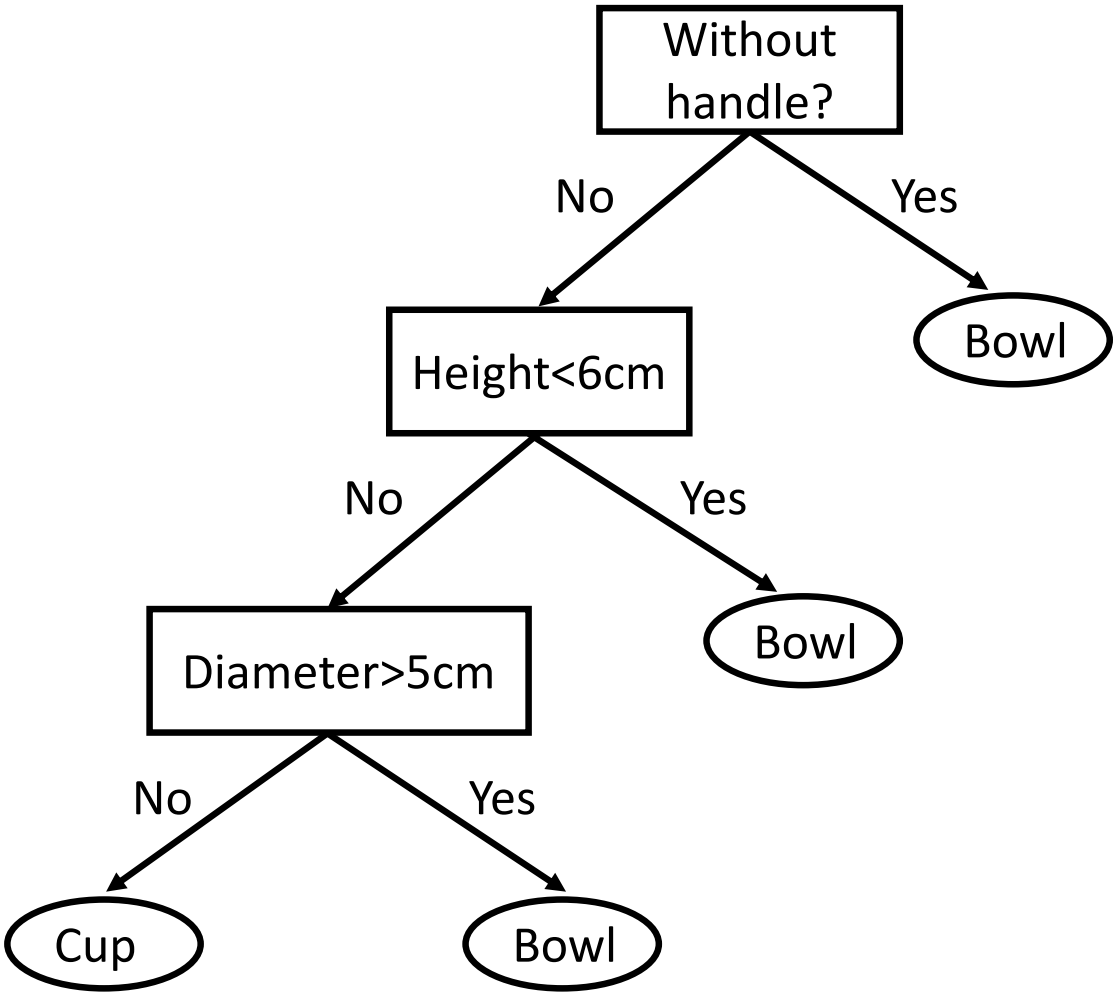

Figure 6.2: A tree showing the distinguishing of Bowl and Cup. Summarizing: The container was a Bowl, if the container (i) without a handle (ii) Height$<$ 6 cm (iii) Diameter $>$ 5 cm.

two more parts at the decision node. The step is continued until some stopping rule is applied (Friedman *et al.*, 2001).

The two critical elements in these steps are the best split strategy and stopping rules.

Split strategy Impurity can be used to evaluate the split strategy of splitting nodes. If one variable at one split point can split samples into two pure parts, then this variable

at this split point can be called the best split strategy at that node.

In our classification trees, we used Gini's diversity index (Loh and Shih, 1997) to compute the impurity of nodes.

The Gini index of a node is:

$$I = 1 - \sum_i p^2(i) \tag{6.12}$$

where $p(i)$ is the class i's fraction of all classes at that node. If a node only has one class, its Gini index is 0. Therefore, the smaller the Gini index (impurity) is, the purer the node is.

The impurity gain (ΔI) between parent and child nodes can be represented as :

$$\Delta I = I(parent) - W_L I(L) - W_R I(R) \tag{6.13}$$

where, $I(parent)$ represents the impurity of the parent node, $I(L)$ represents the impurity of the left child node, and $I(R)$ represents the impurity of the right child node. W_L and W_R are weights of $I(L)$ and $I(R)$. The weight can be calculated by

$$W = \frac{N_c}{N_p} \tag{6.14}$$

N_c is the number of samples at the child node, and N_p is the number of samples at the parent node.

Therefore, the split was chosen through maximising impurity gain.

For regression trees, the mean-square error (MSE) of the responses replaces the Gini index to compute the impurity of nodes. Other steps are the same as for classification trees.

Stopping rules Decision trees need to know when to stop splitting. If there is no constraint on the decision tree, the splitting of the tree will continue until there is only one class (or only one value) at the node. This makes the decision tree huge and causes overfitting problems. This is also time-consuming.

Therefore, some stopping rules are set (The MathWorks, Inc., 2018b):

- Maximal number of decision splits (Max-Splits): It sets the maximal number of decision splits of a tree.
- Minimum number of samples per leaf (Min-Leaf-Size): It sets the minimum number of samples of a leaf node.

- Minimum number of samples per decision node (Min-Dec-Size): It sets the minimum number of samples of a decision node. If both the Min-Leaf-Size and Min-Dec-Size values are set, the decision tree uses the setting that gives larger leaves: Min-Dec-Size=max(Min-Leaf-Size, 2× Min-Leaf-Size).

- Criteria value of impurity: The Gini index (impurity) of 0 is the best value of a node. We also can set a minimum criteria value for impurity which is greater than 0.

The construction of a tree can stop when one of the above rules meets the threshold.

6.3.3 Number and Layers of Weak Learners

In our research, the decision tree was used as weak learner, and the number of weak learners in each ensemble was 100.

For boosting ensemble, "the weak learning method only need to be better than random" (Freund, Schapire, and Abe, 1999) and overly complex weak hypotheses will lead to bad performance (Freund *et al.*, 1999). Therefore, the decision trees are set as stumps, which means the depth of trees is only one layer.

For random forest, the performance depends on the strength of the individual trees and correlation of any two trees (Breiman, 2001). The stronger (lower error rate) the individual trees are, the performance of random forest is better (Breiman and Cutler, 2019). To increase the correlations between trees lead to increasing the random forest's error (Breiman and Cutler, 2019). Therefore, deep trees were built to increase the strength of the individual trees and decrease correlations of trees. Three factors (introduced in Section 6.3.2.2) were set to control the depth of trees:

- Max-Splits: $n - 1$, n is the number of training samples;

- Min-Leaf-Size: 1;

- Min-Dec-Size : 10.

These settings tend to grow deep trees in our models.

6.3.4 Feature Importance

We fed many features (independent variables) into ensemble aggregation algorithms to establish depressivity prediction models, but only a few of them could essentially

influence the final models. These important features are potential biomarkers that we tried to find in EEG.

The relevant importance of each variable X_n in a decision tree can be estimated by summing impurity gains (introduced in Section 6.3.2.2) at nodes which are split by the variable X_n .

The importance of each variable X_n in LSBoost or AdaBoost can be estimated by summing all estimates over all decision trees in the ensemble. The highest value indicates that this feature is the most important one (Friedman *et al.*, 2001; The MathWorks, Inc., 2018b).

For RF and for one of the trees in the forest, out-of-bag samples those were not used to build the tree can be used to calculate the importance of each variable X_n.

- Step1: After each tree (t) is constructed, the prediction error of out-of-bag samples (Err_t) is calculated;
- Step2: The values of the nth feature in the out-of-bag samples are randomly permuted and the permuted out-of-bag samples are fed into the constructed tree to calculate prediction error (permErr_t);
- Step3: Calculates the difference of two prediction errors by Equation 6.15:

$$\Delta\text{Err\_}X_{n,t} = \text{permErr}_t - \text{Err}_t \tag{6.15}$$

- Step4: Repeat step 2 and step 3 for all features;

Then, $\Delta\text{Err\_}X_n$ is averaged over the entire ensemble and divided by the standard deviation over the entire ensemble. If feature X_n is unimportant, after permuting, the prediction error will not change much. Therefore, the highest $\Delta\text{Err\_}X_n$ indicates that this feature is the most important one (Breiman, 2001; The MathWorks, Inc., 2018c).

6.4 Least Absolute Shrinkage and Selection Operator (Lasso)

Least absolute shrinkage and selection operator (Lasso) algorithm was proposed by Tibshirani (1996) to enhance the performance of least squares regression model. The key method of lasso is applying a shrinking process to penalize the coefficients of the regression variables. In this procedure, some variables' coefficients are shrunk to zero.

Therefore, it can "reduce variance without a substantial increase of the bias" (Fonti and Belitser, 2017). The most important advantage of lasso is the same as ensemble learning methods that it can easily select the most important variables. After shrinking, the variables of non-zero coefficients are used to establish regression model, and these variables are important variables.

The lasso is defined by Equation 6.16 (Tibshirani, 1996):

$$\begin{aligned} \hat{\beta} = \text{argmin}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2\} \\ \text{subject to } \sum_{j=1}^{p}|\beta_j| \leq t. \end{aligned} \tag{6.16}$$

where, t is the limit for the sum of the coefficients. Compare to ordinary least square regression, lasso has a restriction shown by the second line in Equation 6.16. The restriction means that the obtained coefficient should be shrunk in a range, controlled by parameter t. This procedure can reduce the model's complexity and shrink coefficients of some variables into zero.

Equation 6.16 is equivalent to the following equation:

$$\hat{\beta} = \text{argmin}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\} \tag{6.17}$$

where, $\lambda(\geq 0)$ is the parameter to control the strength of the penalty. λ is inversely related to t. When $\lambda = 0$, the regression model becomes the ordinary least square regression. Increasing the value of λ can increase the amount of shrinkage (Fonti and Belitser, 2017) and obtain a simpler model.

Overall, lasso can improve the over-fit problem which usually exits in ordinary least square regression method, and it has the ability to select important variables, therefore, in this research, lasso as a regression method is also used to establish depressivity prediction models.

In the implementation, coordinate descent was used as the algorithm to fit the coefficients value. The value of λ was selected by cross-validation. The searching area of λ was defined as $(\lambda_{min}, \lambda_{max})$. The λ_{max} is defined by following equation (Friedman, Hastie, and Tibshirani, 2010):

$$\lambda_{max} = \frac{\max|< x_j, y >|}{N} \qquad j = 1, ..., p \tag{6.18}$$

where, $< x_j, y >= \sum_{i=1}^{n} x_{ij}y_i$. If $\lambda > \lambda_{max}$, $\beta = 0$ (Friedman *et al.*, 2010).

The λ_{min} is set as $\epsilon\lambda_{max}$. Then the searching range was equally divided into D points on the logarithmic scale. Friedman *et al.* (2010) mentioned that the typical value of ϵ is 0.001 and D is 100.

6.5 Conclusion

This chapter discussed the predictive methods we used to establish the depressivity prediction model, three ensemble learning methods, one neural network method and one linear regression method. Their characteristics and algorithms were introduced. One of the ensemble learning methods called LSBoost is suitable to solve regression problems, AdaBoost is suitable to solve classification problems, and random forest is suitable to solve both regression and classification problems. The LVQ neural network is suitable to solve classification problems. The Lasso is suitable to solve regression problems. In Chapters 7 and 8, these methods are applied to the corresponding systems.

Chapter 7

Classification-Based Depressivity Prediction Models

This chapter introduces classification models based on the idea that the occurrence of some specific EEG patterns, which indicate High depressivity, increases in number with a higher depressivity.

The aim of this research is not only to design prediction models to precisely predict depressivity, but also to find potential biomarkers of depressivity in the EEG signal. While establishing models, the most important features to predict depressivity were also determined.

7.1 Introduction

In this chapter, four sets of depressivity prediction models are designed to extract the relevant information from EEG recordings and to predict depressivity based on the information extracted.

In these models, four methods were used to extract features. The first two were classic time-frequency methods: FFT and CWT, which were also used in regression-based depressivity detection models. FFT is a classic method to transfer signals from the time domain to the frequency domain. However, further exploration was required to determine whether the time domain information influences the results. Therefore, the CWT was also used to generate prediction models. Another feature extraction method was the common spatial pattern (CSP) algorithm, which is a spatial filtering technique. CSP is good at distinguishing two different kinds of signals. Therefore, in this chapter, CSP was used to classify high and low depressivity EEG epochs. CSP has

an important advantage that the patterns it generates from EEG can be regarded as a source distribution. This means that we can localize the most related EEG sources based on the patterns and this source distribution can be treated as a biomarker. The fourth feature extraction method was a new method that was a combination of FFT and CSP. The raw data were first transformed by FFT and converted to the power spectrum to obtain EEG power in the frequency domain. Then, the CSP algorithm was applied to the EEG power values to find the most related information of each class (High depressivity or Low depressivity). Those features were then used as inputs into AdaBoost, RF or LVQ for classification. AdaBoost and learning vector quantization (LVQ) are classification algorithms. Random forest (RF) can be used for both regression and classification problems. Therefore, AdaBoost, RF, and LVQ were used separately in classification-based models to classify High depressivity and Low depressivity EEG signals. It should be noted that in this research, the result obtained by the LVQ method was used as a baseline because LVQ is frequently used for the classification of EEG data in previous research (Poulos *et al.*, 1999; Vuckovic *et al.*, 2002; Poulos *et al.*, 2002; Ocak, 2008; Dhar *et al.*, 2017; Pradhan *et al.*, 1996). There was no specific expectation of its performance.

Each feature extraction method was used with all three classification methods. Therefore, the models in this chapter are sorted by feature extraction methods (see Table 7.1). The first set of three models use fast Fourier transform with: AdaBoost (*Cla-FFT-AdaBoost*); random forest (*Cla-FFT-RF*); and learning vector quantization (*Cla-FFT-LVQ*). The second set of three models use continuous wavelet transform with: AdaBoost (*Cla-CWT-AdaBoost*); random forest (*Cla-CWT-RF*); and learning vector quantization (*Cla-CWT-LVQ*). The third set of three models use common spatial pattern with: AdaBoost (*Cla-CSP-AdaBoost*); random forest (*Cla-CSP-RF*); and learning vector quantization (*Cla-CSP-LVQ*); the fourth set of three models use FFT combined CSP technique with: AdaBoost (*Cla-FFT-CSP-AdaBoost*); random forest (*Cla-FFT-CSP-RF*); and learning vector quantization (*Cla-FFT-CSP-LVQ*).

The proposed methods were tested on the depressivity EEG dataset (introduced in Chapter 3). The EEG dataset includes eyes open and closed data epochs. The performances of different feature extraction methods, classification methods, and eye states were compared in each set of models using statistical analysis (mainly using ANOVA). All models' performance were also compared. These comparisons can give us: (i) a hint of how to choose suitable feature extraction methods and predictive analysis methods to predict depressivity using EEG data; (ii) a direction for designing

	FFT	CWT	CSP	FFT+CSP
AdaBoost	Cla-FFT-AdaBoost	Cla-CWT-AdaBoost	Cla-CSP-AdaBoost	Cla-FFT-CSP-AdaBoost
RF	Cla-FFT-RF	Cla- CWT -RF	Cla-CSP-RF	Cla-FFT-CSP-RF
LVQ	Cla-FFT-LVQ	Cla- CWT - LVQ	Cla-CSP-LVQ	Cla-FFT-CSP-LVQ

Table 7.1: Classification-based depressivity prediction models with different feature extraction and predictive methods.

EEG collecting experiments if we know which eyes state can produce better results.

7.2 Overview

The dataset used in this research includes a PID5 depressivity (PID5-d) score for each participant. Therefore, what we need to solve was a regression problem. However, we also designed a series of classification-based models to estimate depressivity because anyone can feel quite depressed briefly, but truly depressed people are likely to feel depressed most of the time. Therefore, we designed classification models based on the assumption that each epoch of signal represents a sign that this epoch is High depressivity or Low depressivity (high or low). If a participant has more High depressivity epochs, it means the participant is highly depressed. If a participant has fewer High depressivity epochs, it means the participant is less depressed. Therefore, in classification-based models, the main aim is to classify the epochs into high or low class but the number of such epochs can then assess the level of depressivity in an individual.

In our dataset, there were only 73 participants (29 males; 44 females), therefore, it was not suitable to split into different gender groups. The reason is explained in Section 3.3.1.1.

Figure 3.3 illustrates the establishment and evaluation steps of depressivity prediction models. They are the participant assignment, feature extraction, predictive analysis, depressivity estimation and model evaluation. Because the aim of classification-based models is to estimate each epoch's class (high or low), in the participant assignment step, we chose high class training data samples, low class training data samples, and testing data samples separately. The details of the participant assignment are discussed in Section 3.3.1.1 of Chapter 3. Next, the features of training samples were

extracted and used to establish prediction models. The testing samples were used to test the models. After each epoch within each testing set has been classified as high or low, the ML-d score of each participant was calculated by Equation 3.2 which is discussed in detail in Section 3.3.3 of Chapter 3. Finally, the effectiveness of the depressivity prediction models was estimated by correlating the PID5-d score against the ML-d score (Section 3.3.4).

Biosignals are complex and sometimes instable. PID5-d scores which are obtained by questionnaire, also have some uncertainty. Therefore, the R^2-values presented in the following sections probably appear low, especially from a computer science perspective. However, in our lab, the same dataset used in this research with an additional 28 participants was tested using traditional EEG alpha asymmetry and HFD (Higuchi's fractal dimension) with stepwise regression (unpublished). The results obtained by the traditional EEG alpha asymmetry method were very low and showed no significant correlation between alpha asymmetry and depressivity. The best R^2-value was obtained by HFD and was only 4%. In a previous study, Stewart and Allen (2018) reported that they examined the relationship between resting frontal EEG asymmetry and BDI scores (BDI is clinical scale for depression), the highest R^2-value obtained by their female group was 11% . Therefore, in the following sections, 11% was set as a threshold to decide if a model is likely to be useful for psychological applications. Furthermore, the corresponding p-value of each R^2-value is also calculated to determine the reliable of the results.

7.3 Classification Using FFT Features (Classification-Based Models I)

In this set of models, FFT was used as feature extraction method. This set of models are *Cla-FFT-AdaBoost*,*Cla-FFT-RF*, and *Cla-FFT-LVQ*. They are collectively known as *Cla-FFTs*.

7.3.1 Feature Extraction

Figure 7.1 shows the data process from raw data to the feature data. On the left of this figure the data preprocessing steps which are introduced in Chapter 3.4 are shown. The right part shows the feature extraction procedure used in this set of models . In this set of models, 2-seconds-length epochs with 1-second overlap were transformed

by FFT and constructed to feature vectors. Section 5.3.1 describes the details of the implementation. Since there were 18 channels per epoch and 159 frequency samples per channel, each epoch was represented as a 2862 dimensional feature vector. Each dimension was considered a feature. Section 5.3.2 describes the feature vector construction method in detail. The average number of epochs extracted per participant for eyes open or eyes closed was 228.99 and the minimum number extracted was 192 (out of a maximum of 230).

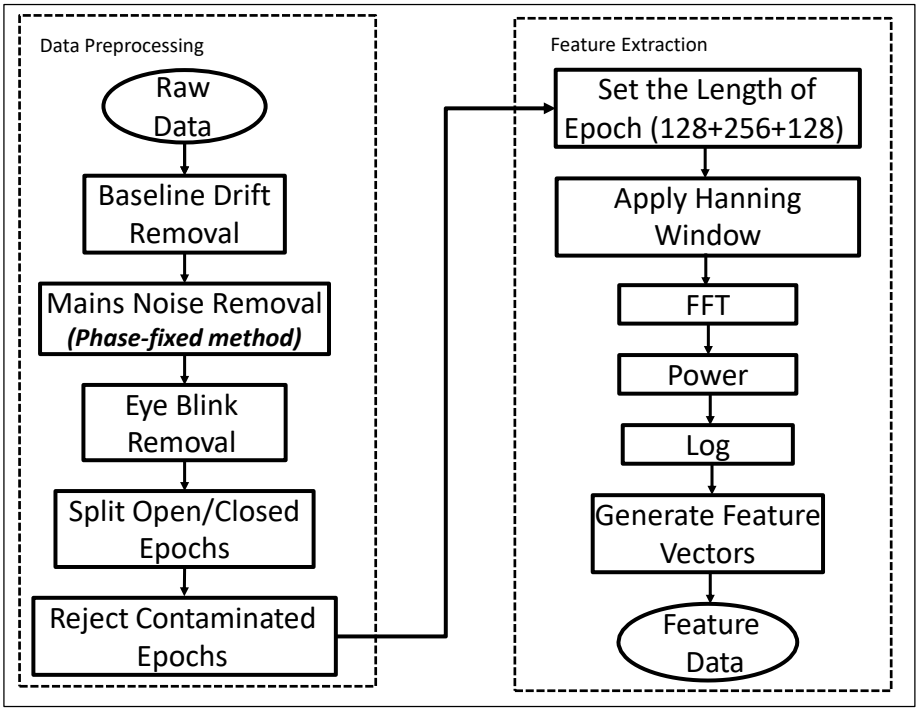

Figure 7.1: This figure shows the data process steps from raw data to the feature data of models *Cla-FFTs*. The models *Reg-FFTs* introduced in the next chapter use the same data process steps.

7.3.2 Predictive Analysis

In AdaBoost and RF, the decision tree was used as the weak learner and the number of weak learners was 100. For AdaBoost, the decision trees are stumps, which means the depth of trees was only one layer. For RF, deep trees were used. The depth of trees were controlled by three factors that were introduced in Section 6.3.3. For RF, The number of features searched at each split point (m) was: $m = \sqrt{p}$ (p is the number of all features). The details are provided in Section 6.3.1.3.

The number of hidden neurons for LVQ was set at 20 which means we assume there were 10 high EEG patterns and 10 low EEG patterns respectively. The error tolerance was set to 0.1. In this project, the EEG data were complex, and it was not so easy to distinguish high and low class EEG. Sometimes, before the algorithm met the error tolerance (0.1), the maximum number of iterations (100) was reached, the training would be stopped.

7.3.3 Experiments and Results

For different eye states, a set of experiments (ABBA-Open, ABBA-Closed, and ABBA-All) were implemented to predict participants' depressivity in each model. In these experiments, we used the ABBABAABBAAB sequence participant assignment method (abbreviated as ABBA) to select training and testing samples (The ABBA sequence participant assignment method is described in Section 3.3.1.1 of Chapter 3). The second set of experiments (TOP-Open, TOP-Closed, and TOP-All) used the top 6 PID5-d score samples and bottom 6 samples as training data samples, and the remaining samples as testing samples. The results of these two sets of experiments were compared. Table 7.2 shows the R^2-values and the corresponding p-values of these experiments by correlating the PID5-d scores against the ML-d scores of each participant in different experiments.

Table 7.2 shows that the two highest R^2-values (10.97%, 8.37%) were obtained by the experiments ABBA-Open and ABBA-All of model *Cla-FFT-RF*. The highest value is approximate 11%, comparable to the results previously reported in paper (Stewart and Allen, 2018). Figure 7.2 shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment ABBA-Open of model *Cla-FFT-RF*.

In Table 7.2, there are three independent variables (factors). They are eye state, participant assignment method, and predictive analysis method. To test the effects of multiple factors in models, the R^2-values of all experiments were submitted to a three-way ANOVA using the highest order interaction as the error terms since there are no replicates. As can be seen in Table 7.2, the ABBA participant assignment method was much more successful than the TOP method (PA, $F(1, 4) = 192.16$, $p < 0.001$). Table 7.2 also shows that different eye states had influence on depressivity predictive ability (Eyes, $F(2, 4) = 8.2$, $p = 0.0384$). The results obtained by eyes closed epochs were worse than eyes open and all eye states. From the results, we can also see that the depressivity prediction ability of LVQ is worse than other two methods (PMethods, $F(2, 4) = 42.3$, $p = 0.002$).

PA	Eyes	*AdaBoost*	*RF*	*LVQ*
		R^2-*value*(*p-value*)	R^2-*value*(*p-value*)	R^2-*value*(*p-value*)
ABBA	Open	7.26% (0.0212)	10.97% (0.0042)	2.61% (0.1724)
	Closed	6.11% (0.0350)	4.79% (0.0628)	2.28% (0.2021)
	All	7.62% (0.0181)	8.37% (0.0130)	1.53% (0.2974)
TOP	Open	0.53% (0.6192)	3.41% (0.2043)	0.30% (0.7098)
	Closed	1.43% (0.4138)	0.22% (0.7494)	0.07% (0.8563)
	All	1.97% (0.3366)	0.80% (0.5413)	0.20% (0.7614)

Table 7.2: **Results of models *Cla-FFTs*.** The first and second columns of the table show the different participant assignment methods for each experiment and the eye states of experiment data. ABBA means using our proposed participant assignment method (Introduced in Section 3.3.1.1 of Chapter 3.) to select training and testing data samples; TOP means choosing the top 6 PID5-d score samples and the bottom 6 samples as training data samples, and the remaining samples as testing samples. Open means only eyes open epochs were used; Closed means only eyes closed epochs were used; and All means both eyes open and closed epochs were used. Values in the table are R^2-values and the corresponding p-values (in brackets) of each experiment. In this set of experiments, R-values were all positive and can be got back by taking the square root of the R^2-values.

7.4 Classification Using CWT Features (Classification-Based Models II)

In this set of models, we used CWT to obtain EEG features. Like the first set of models, AdaBoost, RF, and LVQ were separately used as classification methods. This set of models are *Cla-CWT-AdaBoost, Cla-CWT-RF*, and *Cla-CWT-LVQ*. They are collectively known as *Cla-CWTs*.

7.4.1 Feature Extraction

Figure 7.3 shows the data preprocessing and feature extraction flow of this set of models. Signals were simply split into basic 1-second-length epochs without overlap. The sample rate of our data is 256 Hz (1s), and the total scale of CWT was set to 1:1:128. The components with frequencies less than 80 Hz were reserved. This resulted in 80 frequency components from 1 to 80 Hz. Finally, the square of each component

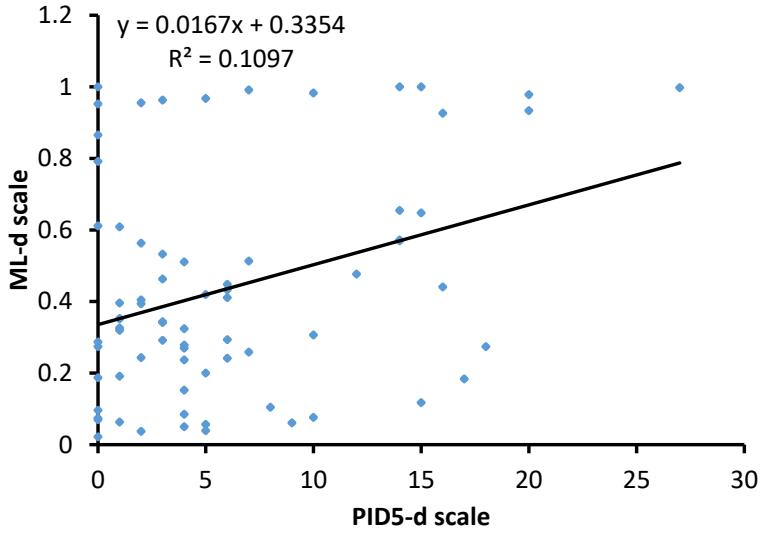

Figure 7.2: The ML-d score and the matched PID5-d score of the experiment ABBA-Open of model *Cla-FFT-RF*.

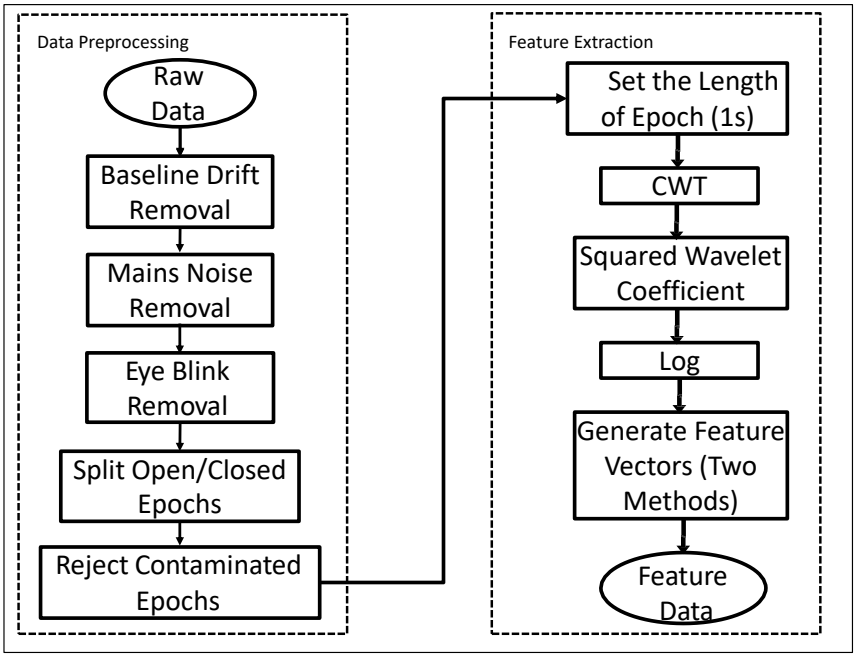

Figure 7.3: This figure shows the steps of the data process from raw data to the feature data of models *Cla-CWTs*.

(power spectrum) was log transformed to normalize error variance and used to construct feature vectors.

In the step of feature vectors construction, we constructed two kinds of feature vectors:

1) Accumulated energy All values along the time domain were averaged to calculate accumulation of energy in each frequency slot.

2) Top five sudden changes The top five maximum squared wavelet coefficient values in each frequency slot were picked to form a feature vector.

The details of CWT and feature vectors construction methods are discussed in Section 5.3.3 and 5.3.4.

7.4.2 Predictive Analysis

The predictive analysis is the same as for the *Cla-FFTs*.

7.4.3 Experiments and Results

As a result of the different eye state epochs and the different feature vector construction (FVC) methods , six experiments were conducted for each model. Table 7.3 shows the R^2-values and the corresponding p-values of these experiments by correlating the PID5-d scores against the ML-d scores of each participant in different experiments.

Table 7.3 shows that the two highest R^2-values (9.29%, 8.45%) were obtained by the experiments Accumulation-Closed of model *Cla-CWT-RF* and model *Cla-CWT-AdaBoost*. The highest value is slightly worse than 11%. Figure 7.4 shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment Accumulation-Closed of model *Cla-CWT-RF*.

Table 7.3 shows the effects of three independent variables (factors): eye state, feature vector construction method, and predictive analysis method. To test the effects of these multiple factors on models, the R^2-values of all experiments were submitted to a three-way ANOVA with the highest order interaction used as the error term (as above).

Unlike the model *Cla-FFTs*, Table 7.3 shows that with the model *Cla-CWTs*, the results of the experiments which used eyes closed epochs are better than other two eye states (Eyes, F (2, 4) = 7.66, p = 0.0429). In this set of models, we can see that other than the depressivity prediction ability of LVQ is still worse than other two methods (PMethods, F (2, 4) = 13.22, p = 0.0173). The result also shows that

FVC	Eyes	*AdaBoost*	*RF*	*LVQ*
		R^2-*value*(*p-value*)	R^2-*value*(*p-value*)	R^2-*value*(*p-value*)
Accumulation	Open	1.50% (0.3014)	4.18% (0.0827)	4.60%(0.0683)
	Closed	8.45% (0.0126)	9.29% (0.0087)	0.44%(0.5758)
	All	3.96% (0.0916)	4.40% (0.0750)	4.05%(0.0879)
Top 5	Open	0.38% (0.6024)	3.03% (0.1407)	4.68%(0.0659)
	Closed	5.73% (0.0413)	6.21% (0.0334)	0.08%(0.8134)
	All	4.16% (0.0836)	5.52% (0.0455)	3.19%(0.1308)

Table 7.3: **Results of models *Cla-CWTs*.** The first and second columns of the table show different feature vector construction methods and different eye states of each experiment. Accumulation means after cwt transformation, all values along the time domain were averaged to obtain one value in each frequency slot; Top 5 means after cwt transformation, the top five maximum values at time domain in each frequency slot were chosen to form a feature vector. Values in the table are R^2-values and the corresponding p-values (in brackets) of each experiment. In this set of experiments, R-values were all positive and can be got back by taking the square root of the R^2-values.

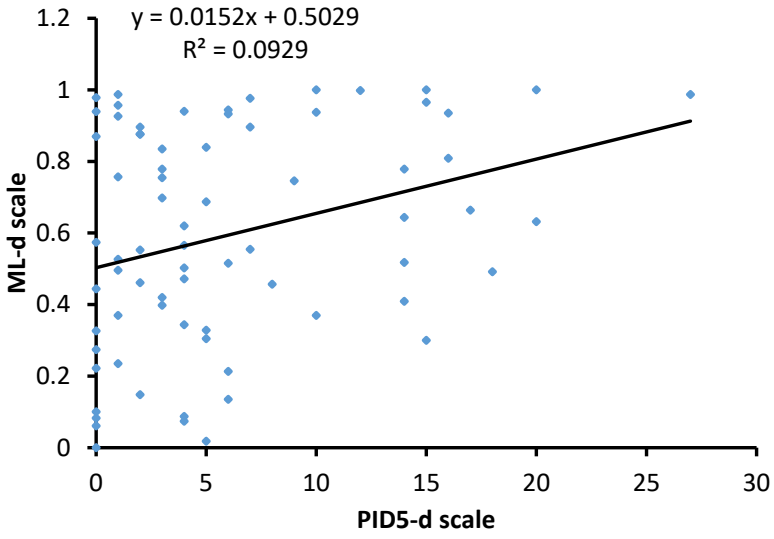

Figure 7.4: The ML-d score and the matched PID5-d score of the experiments Accumulation-Closed of model *Cla-CWT-RF*.

although Accumulation feature vector construction methods is better than Top 5, but there is only a small gap (FVC, F (1, 4) = 4.5, p = 0.1012).

7.5 Classification Using CSP Features (Classification-Based Models III)

In this set of classification-based models, to further explore feature extraction in EEG signal, common spatial pattern (CSP) (Blankertz *et al.*, 2008) was used. This set of models are *Cla-CSP-AdaBoost*, *Cla-CSP-RF*, and *Cla-CSP-LVQ*. They are collectively known as *Cla-CSPs*.

7.5.1 Feature Extraction

Figure 7.5 shows the feature extraction steps of model *Cla-CSPs*. The data preprocessing steps are the same as the previous two sets of models, so they are not repeated here. Different from FFT and CWT, the implementation steps of CSP in the training phase and testing phase is not the same. Figure 7.5 shows that each data sample was split into 1-second-length epochs. In the training phase, the epochs of two training groups were pooled separately, then they were used to generate a CSP filter (P^T in Equation 5.15) through the CSP algorithm. After that, each epoch of all data samples was filtered by the CSP filter to obtain a feature matrix (Z^i in Equation 5.15). The first n rows of the feature matrix contain the most relevant information of the first class; the last n rows of the feature matrix contain most relevant information of the second class; and the middles rows have less information for both classes. However, we do not know the exact number of rows that contain the most information. Therefore, in this research, we set n from 1 to 5 and to find the most suitable number of n to choose.

In the step of feature vectors construction, we constructed two kinds of feature vectors:

1) Whole feature rows: The selected rows of feature matrix were put together to align a feature vector.

2) Variance of rows: The variances of selected rows of feature matrix were calculated to construct a feature vector (see Equation 5.17).

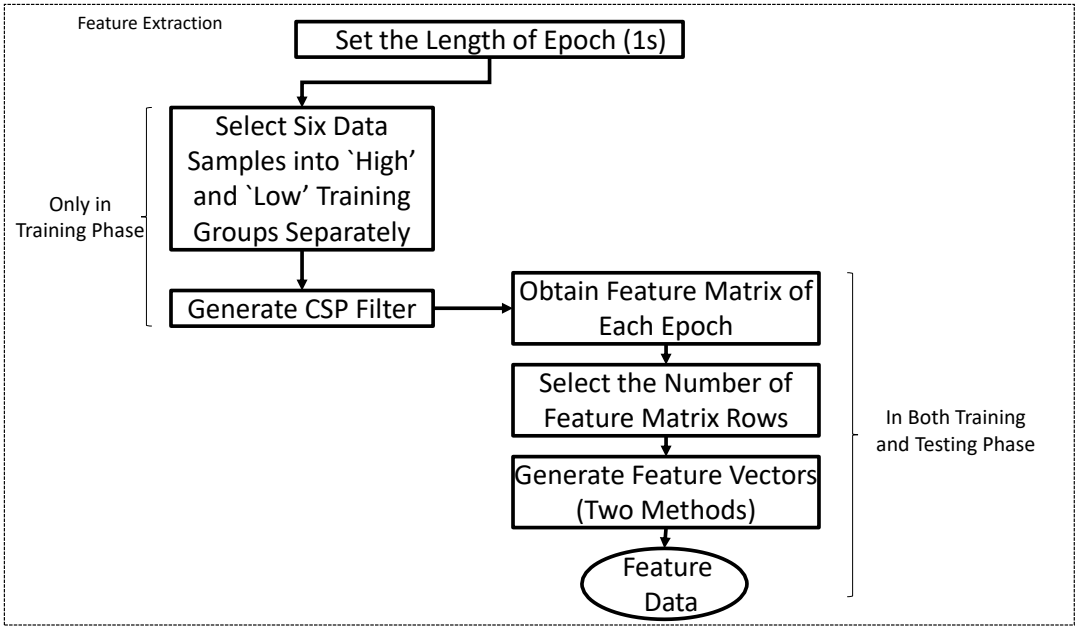

Figure 7.5: This figure shows the feature extraction steps of model *Cla-CSPs*. Each data sample was split into 1-second epochs. In the training phase, the epochs of high training groups and low training groups were pooled separately to generate a CSP filter through the CSP algorithm. Then, each epoch of all data samples was filtered by the CSP filter to obtain a feature matrix. Because the first and last n rows of the feature matrix contain the most relevant information of each class, these $2 \times n$ rows were chosen to construct feature vectors. These vectors are the feature data to train and test models. In this research, n varies from 1 to 5.

The details of CSP and feature vectors construction methods are discussed in Sections 5.3.5 and 5.3.6.

7.5.2 Predictive Analysis

The predictive analysis is the same as models *Cla-FFTs*.

7.5.3 Experiments and Results

For different eye states, feature vector construction methods, predictive analysis methods, and rows of feature matrix, 90 (3 eyes × 2 FVC × 3 Pmethods × 5 n) experiments were conducted to predict participants' depressivity in each model.

7.5.3.1 The number of pairs of feature matrix rows

As mentioned in Chapter 5, a parameter that we can change in CSP is the number of pairs of feature matrix rows. In this section, the influence of the number of pairs of feature matrix rows were assessed. Table 7.4 lists the R^2-values using different numbers of feature matrix rows: from 1 pair to 5 pairs (n = 1 to n = 5) obtained by different models. These values were averaged from different eye states.

n	AdaBoost		RF		LVQ	
	Whole	Var	Whole	Var	Whole	Var
2×1	6.52%	2.89%	6.63%	4.56%	4.37%	3.79%
2×2	8.44%	8.07%	11.66%	5.21%	2.99%	5.76%
2×3	10.90%	5.67%	13.07%	6.09%	2.53%	6.94%
2×4	11.34%	8.21%	13.18%	6.28%	3.86%	6.29%
2×5	11.38%	9.58%	13.47%	9.25%	5.53%	6.08%

Table 7.4: **Results of different numbers of feature matrix rows of models *Cla-CSPs*.** The first column shows the number of pairs of rows used to construct feature vector. Whole and Var are two different feature vector construction methods. Values in the table are R^2-values were averaged from different eye states of each experiment.

The results show that although the first pair of feature matrix row includes the most related information of two kinds of signals, but when we increase n (using more pairs of feature matrix rows), the R^2-values increase. In most cases, the values when n = 5 are best (n, F (4, 8) = 5.93, p = 0.0162). Therefore, in the following section, only the results which were obtained when n = 5 are listed.

The results also show that the LVQ method is worse than other predictive method (Pmethods, F (2, 8) = 17.95, p = 0.0011); and that the results obtained with the Var feature vector are worse than with Whole (Pmethods, F (1, 8) = 11.8, p = 0.0089).

7.5.3.2 General method

Table 7.5 shows the results of the experiments obtained with model *Cla-CSPs*.

The results of Table 7.5 show that the two highest R^2-values (17.80%, 17.47%) were obtained by experiment Whole-All and Var-All of model *Cla-CSP-RF*. These two results are much better than 11% and the results obtained by previous two sets of

FVC	Eyes	*AdaBoost*	*RF*	*LVQ*
		R^2-*value*(*p-value*)	R^2-*value*(*p-value*)	R^2-*value*(*p-value*)
Whole	Open	15.42% (0.0006)	16.15% (0.0004)	15.72% (0.0005)
	Closed	2.78% (0.1586)	6.45% (0.0301)	0.38% (0.6054)
	All	15.94% (0.0005)	17.80% (0.0002)	0.49% (0.5543)
Var	Open	11.71% (0.0030)	10.10% (0.0061)	9.34% (0.0086)
	Closed	0.15% (0.7418)	0.18% (0.7198)	-0.19% (0.7153)
	All	16.88% (0.0003)	17.47% (0.0002)	9.10% (0.0095)

Table 7.5: **Results of models *Cla-CSPs*** The first and second columns of the table shows the different feature vector construction methods and different eye states of each experiment. Whole means the feature vector was obtained by whole rows of the feature matrix; Var means the feature vector was obtained by calculating the variances of rows of the feature matrix. The details are described in Section 5.3.6. As mentioned above, we tried $n \in [1, 5]$, but this table only lists the results when $n = 5$. In most cases, this led to the best performance. Values in the table are R^2-values and the corresponding p-values (in brackets) of each experiment. The negative R^2-value represents the corresponding *R-value* was negative.

models. Figure 7.6 shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment Whole-All of model *Cla-CSP-RF*.

Table 7.5 shows the effects of three independent variables (factors): eye state, feature vector construction method, and predictive analysis method. To test the effects of these multiple factors on models, the R^2-values of all experiments (Table 7.5) were submitted to a three-way ANOVA with the highest order interaction used as the error term (as above).

Table 7.5 shows that R^2-values obtained with eyes open epochs and all epochs are higher than eyes closed epochs (Eyes, F (2, 4) = 61.81, p = 0.001). The R^2-values obtained with different feature vector construction methods do not differ (FVC, F (1, 4) = 3.56, p = 0.1323). The performance of AdaBoost and RF are similar, but they are much better than LVQ (PMethod, F (2, 4) = 12.74, p = 0.0184).

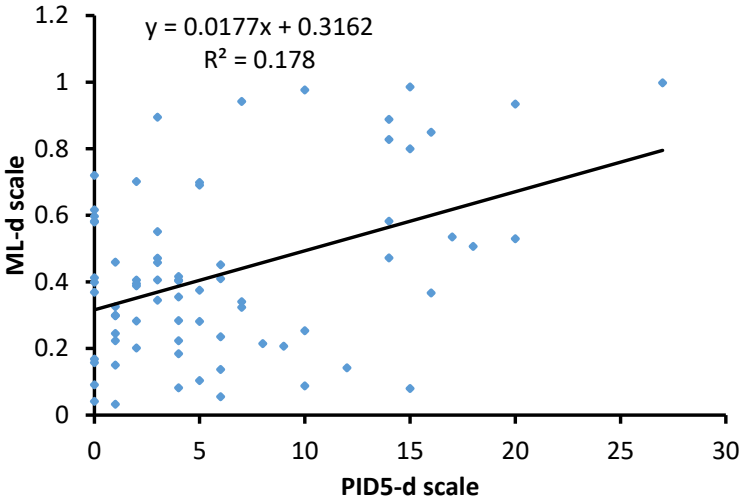

Figure 7.6: The ML-d score and the matched PID5-d score of the experiment Whole-All of model *Cla-CSP-RF*.

7.6 Classification Using FFT Combining CSP Features (Classification-Based Models IV)

7.6.1 Introduction

In the previous set of models, we used the CSP algorithm to extract spatial features from EEG data to predict depressivity. In the fourth set of classification-based models, we designed a new method where the raw data were first transferred by FFT to obtain the log of EEG power values in the frequency domain. Then, the values were filtered via CSP to obtain the most depressivity-related EEG components.

This set of models are *Cla-FFT-CSP-AdaBoost*, *Cla-FFT-CSP-RF*, and *Cla-FFT-CSP-LVQ*. They are collectively known as *Cla-FFT-CSPs*.

7.6.2 Feature Extraction

Figure 7.7 illustrates the details of feature extraction steps in this set of models. Each epoch of raw data ($N \times T$, where N is the number of channels, and T is the number of time points) was first transformed with the FFT method. The implementation of FFT was the same as the feature extraction methods of the models *Cla-FFTs* (see Section

7.3.1). After that, the new epoch of data were formatted as $N \times F$ (N is the number of channels, and F is the number of frequency points). Then, the $N \times F$ matrix was fed into the CSP algorithm to obtain the features. This last part is the same as the feature extraction methods of models *Cla-CSPs* (see Section 7.5.1).

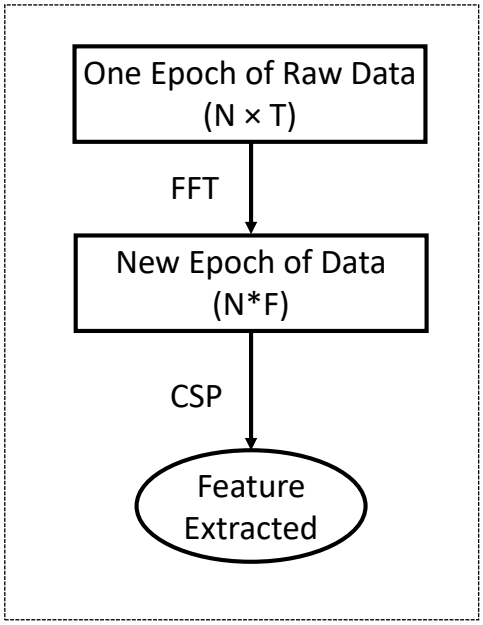

Figure 7.7: This figure illustrates the feature extraction details of models *Cla-FFT-CSPs*. Each epoch of raw data ($N \times T$, where N is the number of channels, and T is the number of time points) was first transferred by the FFT method. The implementation of FFT is the same as the feature extraction methods of the models *Cla-FFTs* (see Section 7.3.1). After that, the new epoch of data was formated as $N \times F$ (N is the number of channels, and F is the number of frequency points). Then, the $N \times F$ matrix was fed into the CSP algorithm to obtain the features. This part is the same as the feature extraction methods of models *Cla-CSPs* (see Section 7.5.1).

7.6.3 Predictive Analysis

The predictive analysis is the same as models *Cla-FFTs*.

7.6.4 Experiments and Results

In this set of models, we made a range of adjustments to the procedure. Before describing the different experiments, the basic parameters are explained here.

The data of each participant were split into 2-seconds-length epochs with 1-second overlap (the method is introduced in Section 5.2). Then each epoch was transformed using FFT to obtain the log of power of the EEG signal. After the FFT, the epochs from the training groups were pooled to calculate the CSP spatial filter and generate the training classifier (using AdaBoost, RF and LVQ separately). The epochs of the remaining participants were used to be tested.

Like model $Cla-CSPs$, in this set of models, we conducted 90 (3 eyes × 2 FVC × 3 Pmethods × 5 n) experiments for different eye states, feature vector construction methods, predictive analysis methods, and rows of feature matrix to predict participants' depressivity in each model.

7.6.4.1 The number of feature matrix rows

In this section, the influence of the number of feature matrix rows was tested. Table 7.6 lists the R^2-values using different numbers of pairs of feature matrix rows: from 2×1 to 2×5 (n = 1 to n = 5). These values were averaged from different eye states.

n	AdaBoost		RF		LVQ	
	Whole	Var	Whole	Var	Whole	Var
2×1	3.36%	3.42%	3.63%	4.56%	3.06%	3.80%
2×2	3.66%	1.92%	3.76%	1.14%	4.62%	0.51%
2×3	1.77%	4.10%	1.63%	2.27%	3.37%	1.20%
2×4	0.88%	1.99%	1.53%	0.33%	3.46%	-0.85%
2×5	0.88%	3.30%	1.46%	1.12%	3.00%	-0.34%

Table 7.6: **Results of different numbers of feature matrix rows of models *Cla-FFT-CSPs*.** The first column shows the number of pairs of rows used to construct feature vector. Whole and Var are two different feature vector construction methods. Values in the table are R^2-values were averaged from different eye states of each experiment.

These results show the opposite results to model *Cla-CSPs*, in most cases, the values in $n \in [2, 5]$ are worse than n = 1 (n, F (4, 8) = 5.99, p = 0.0157). Therefore,

in the following section, only the results which were obtained when n = 1 are listed. The results also suggest that the results obtained by Whole feature vector construction method are slightly better than the results obtained by Var, but the difference is not significant (FVC, F (1, 8) = 4.99, p = 0.0559). As can be seen from the results that there are no big difference between the three prediction methods (Pmethods, F (2, 8) = 0.5, p = 0.6244).

7.6.4.2 General method

Table 7.7 shows the results of the experiments obtained by model *Cla-FFT-CSPs*. As mentioned above, we tried $n \in [1, 5]$, but this table only lists the results when n = 1 . In most cases, this led to the best performance.

FVC	Eyes	*AdaBoost*	*RF*	*LVQ*
		R^2-*value*(*p-value*)	R^2-*value*(*p-value*)	R^2-*value*(*p-value*)
Whole	Open	5.21% (0.0521)	5.59% (0.0434)	0.96%(0.4104)
	Closed	2.37% (0.1938)	2.23% (0.2076)	3.45%(0.1154)
	All	2.51% (0.1810)	3.08% (0.1374)	4.77%(0.0634)
Var	Open	8.63% (0.0116)	11.07% (0.0040)	9.99% (0.0065)
	Closed	-0.07%(0.8186)	-0.09% (0.8017)	-0.22% (0.6912)
	All	1.70% (0.2714)	2.71% (0.1637)	1.64% (0.2803)

Table 7.7: **Results of models *Cla-FFT-CSPs*.** The first column of table shows different feature vector construction methods used in experiments. This table only lists the results when n = 1 (the definition of n is shown in Section 7.5.1). Values in the table are R^2-values and the corresponding p-values (in brackets) of each experiment. The negative R^2-values represent the corresponding R-values were negative.

Table 7.7 shows that the two highest R^2-values (11.07%, 9.99%) were obtained by the experiments Var-Open of model *Cla-FFT-CSP-RF* and model *Cla-FFT-CSP-LVQ*. The highest result is worse than the highest result obtained by *Cla-CSPs*, while it is slightly over 11%. Figure 7.8 shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment Var-Open of model *Cla-FFT-CSP-RF*.

Table 7.7 shows the effects of three independent variables (factors): eye state, feature vector construction method, and predictive analysis method. To test the effects of these multiple factors on models, the R^2-values of all experiments were submitted to a three-way ANOVA with the highest order interaction used as the error term (as

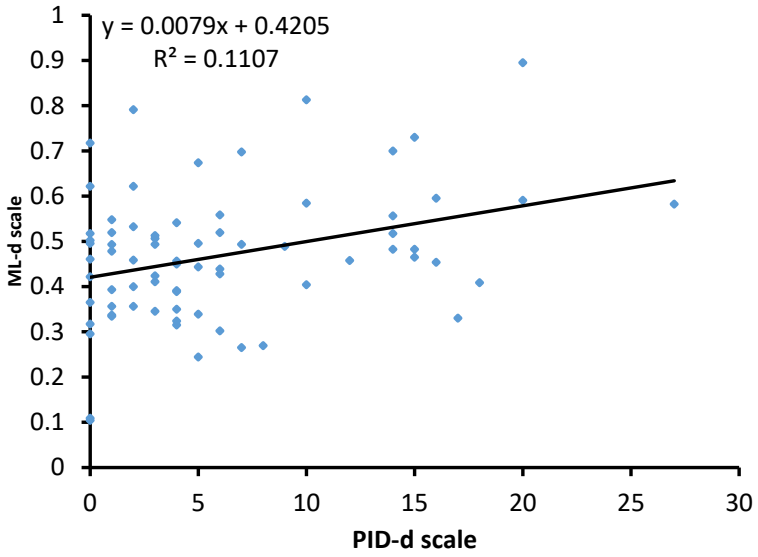

Figure 7.8: The ML-d score and the matched PID5-d score of the experiments Var-Open of model *Cla-FFT-CSP-RF*.

above).

Table 7.7 shows that with the model *Cla-FFT-CSPs*, the results of the experiments which used eyes closed epochs are worse than other two eye states (Eyes, F (2, 4) = 20.11, p = 0.0082). But the results do not show big difference between two feature vector construction methods (FVC, F (1, 4) = 0.59, p = 0.4865) or three predictive analysis methods (PMethods, F (2, 4) = 0.37, p = 0.7105).

7.7 Important Features

In the process of establishing depressivity prediction models, the important features which were most related to depressivity were discovered by AdaBoost, RF, and CSP. In this section, the important features obtained by good behaviour models are shown.

Figure 7.9 shows the results of classification models designed in this chapter. The results obtained by LVQ are almost all worse than AdaBoost and RF, therefore, the results obtained by LVQ were not plotted.

The results marked by blue diamonds were obtained by using eyes open epochs; the results marked by red hollow circles were obtained by using eyes closed epochs; the results marked by black asterisks were obtained by using all eye states epochs. We can

see that the overall performance of eyes closed epochs is worse than eyes open epochs or all eye states epochs, except for model *Cla-CWTs*. The performance of eyes-Open and all eye states are similar. The R^2-values obtained by model *Cla-CSP-RF* using two feature vector construction methods with all eye states epochs are all over 17%. These results are the best overall and they are much better than the results (11%) reported by paper (Stewart and Allen, 2018). Therefore, the important features obtained by model *Cla-CSP-RF* with corresponding eye states are described here and will be compared with the features obtained by regression-based models to see if the important features are consistent.

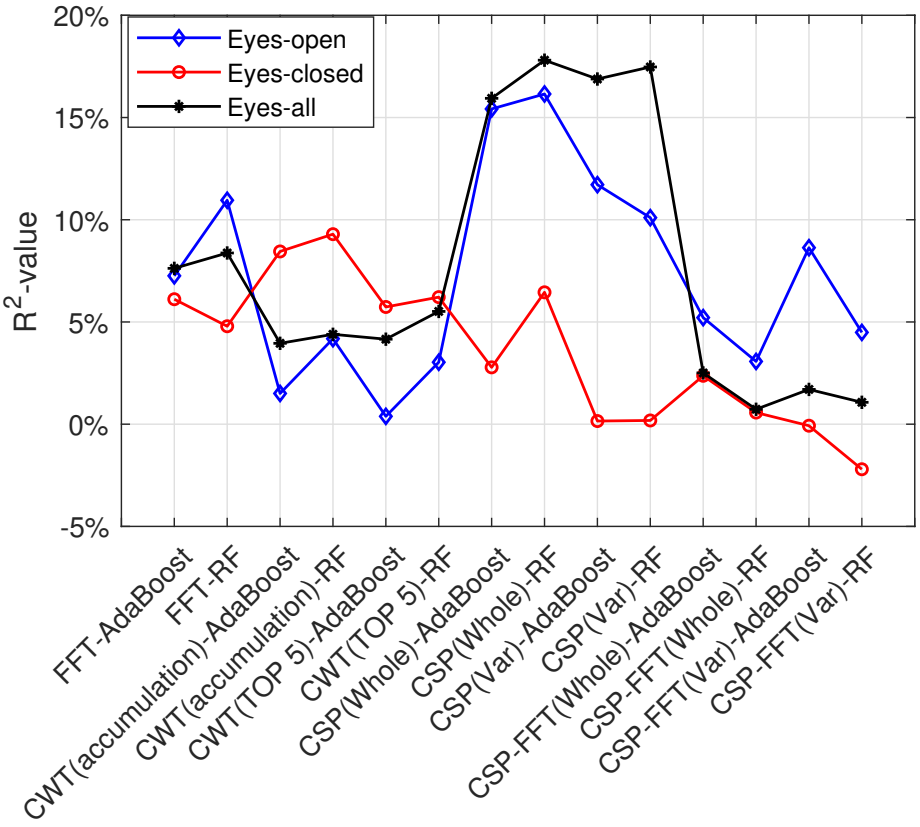

Figure 7.9: This figure shows the results of classification-based models designed in this chapter. The results marked by blue diamonds were obtained by using eyes open epochs; the results marked by red hollow circles were obtained by using eyes closed epochs; the results marked by black asterisks were obtained by using all eye states epochs. The corresponding models used to obtain these results are listed in the horizontal axis.

In model *Cla-CSPs*, the common spatial pattern algorithm is used to extract features from EEG signal. CSP itself can supply the spatial distribution from the two kinds of EEG signals. Because the performance of using all eyes epochs is the best of all results of this model, the common spatial pattern obtained by this group is plotted. The patterns were calculated by Equation 5.16. We implemented experiments with ABBA sequence participant assignment method. In each experiment, we established 2 ensembles in two rounds (round A and round B). Therefore, we merged the two sets of patterns obtained from round A and round B by averaging these two sets of patterns. These two sets of patterns were changed into absolute value and normalized into $(0, 1)$ in advance. Figure 7.10 shows the first two important spatial patterns of High depressivity. From these two spatial patterns, we can see that the optimal channels for detecting High depressivity epochs were T7, FP2, T8 and C3.

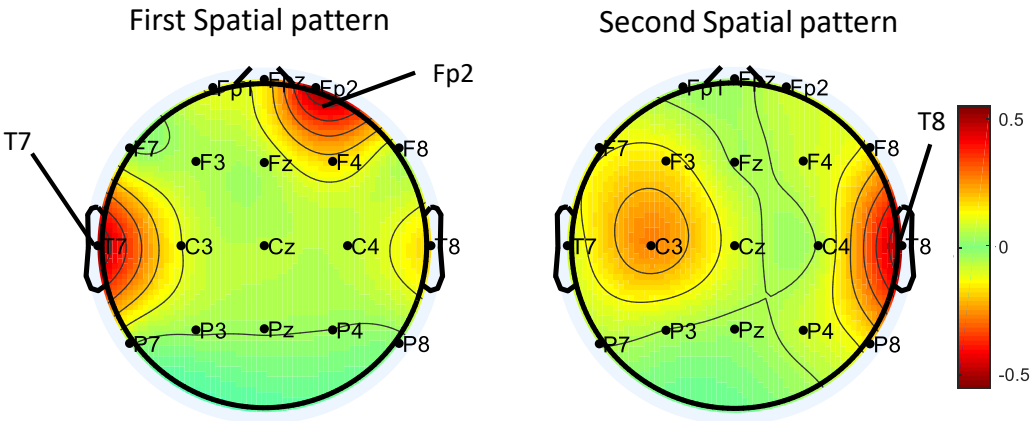

Figure 7.10: This figure shows the first two important CSP patterns of High depressivity obtained by using all eye states epochs. From the two spatial patterns, we can see that the optimal channels for detecting High depressivity epochs are T7, FP2, T8 and C3.

The other important features with statistically significant R^2-values (p-value $<= 0.05$) obtained in the rest of the experiments are listed in Table 7.8. In the table, the important features marked with ** were obtained by experiments which obtained very reliable R^2-values (p-value $<= 0.001$). Optimal channels for detecting High depressivity epochs were T7, FP2, T8 and C3. These channels were obtained from the model *Cla-CSPs* with all eye states epochs. Another set of important features was obtained from

the model *Cla-CSPs* with eyes open epochs (Table 7.8). The two important channels are again Fp2 and T7. The remaining important features listed in Table 7.8 are almost all focused on the T7, T8, and C3, and the *gamma* band. Overall, T7, T8, and C3 appear important for prediction of depressivity; Fp2 might be the key factor to help the model *Cla-CSPs* achieve high depressivity prediction performance.

Eyes	Open		Closed		All	
Feature / Model	Channel	Freq (Hz)	Channel	Freq (Hz)	Channel	Freq (Hz)
Cla-FFT-Adaboost	T8, C3	50, 50.5	C3, T8	50, 11	T8, C3	50, 49.5
Cla-FFT-RF	T8, T7	50, 50.5	/	/	T7, T8	50, 11
Cla-CWT(A)-Adaboost	/	/	T7, C3	71, 73	/	/
Cla-CWT(A)-RF	/	/	T7, C3	4, 80	/	/
Cla-CWT(T)-Adaboost	/	/	T7, P8	78, 76	/	/
Cla-CWT(T)-RF	/	/	T7, T8	79, 80	T7, C3	78, 77
Cla-CSP	Fp2, T7**	/	/	/	See Fig. 7.10**	/
Cla-CSP-FFT	T8, T7	/	/	/	/	/

** p-value <=0.001

Table 7.8: Important features obtained from the classification-based models. All the features listed in this table were all obtained from the experiments with statistically significant R^2-values (p-value <= 0.05). The CWT(A) in the table means CWT(Accumulation); CWT(T) means CWT(Top 5); Freq (Hz) means Frequency (Hz).

7.8 Conclusion

In this chapter, four sets of classification-based models were designed to predict depressivity using EEG data and a series of experiments were implemented to test the performance of these models.

Figure 7.9 shows that the highest R^2-values of these experiments were all obtained by model *Cla-CSPs* with all eye states epochs or eyes open epochs. The highest R^2-value is 17.8%. Compared to the result reported by Stewart and Allen (2018), where the highest R^2-value obtained by female group was only 11%, ours is noticeably better.

The summary table 7.8 shows that the optimal channels and frequency band for detecting High depressivity are T7, T8 and C3 and *gamma* band. These locations

and frequency band are probably important to assess the depressivity of participants. Fp2 might be the key factor to help the model *Cla-CSPs* achieve high depressivity prediction performance.

To further increase the predicting ability and find more precise information about depressivity, in the following chapter, two sets of regression-based depressivity prediction models were designed.

As we need to compare the results obtained from both Chapter 7 and 8, all the results and related work will be discussed in the last chapter (Chapter 9).

Chapter 8

Regression-Based Depressivity Prediction Models

In Chapter 7, we developed four sets of classification-based depressivity prediction models. In this chapter, we introduce two sets of regression models based on the idea that the EEG patterns vary smoothly with the level of depression (depressivity) to predict the depressivity based on their EEG data.

8.1 Introduction

In this chapter, two sets of depressivity prediction models are designed to predict depressivity based on EEG recordings. In these two sets of models, FFT and CWT techniques were employed to extract the EEG power information from the raw data. Then, the LSBoost, RF, and lasso were implemented on the feature vectors to predict the depressivity. Because CSP and LVQ are only suitable for classification, they were not used to design regression-based models. In the procedure of predicting depressivity, the important features which were most related to depressivity were also discovered by LSBoost, RF, and lasso. The important features obtained by well behaved models can be considered as potential biomarkers.

The two feature extraction methods were separately used with all three regression methods. Therefore, the models in this chapter are sorted by feature extraction method (see Table 8.1). The first set of three models use fast Fourier transform with: LSBoost (*Reg-FFT-LSBoost*); random forest (*Reg-FFT-RF*); and Least Absolute Shrinkage and Selection Operator (*Reg-FFT-Lasso*). The second set of three models use continuous wavelet transform with: LSBoost (*Reg-CWT-LSBoost*); random forest (*Reg-CWT-RF*);

and Least Absolute Shrinkage and Selection Operator (*Reg-CWT-Lasso*).

	FFT	CWT
LSBoost	Reg-FFT-LSBoost	Reg-CWT-LSBoost
RF	Reg-FFT-RF	Reg- CWT -RF
Lasso	Reg-FFT-Lasso	Reg- CWT - Lasso

Table 8.1: Regression-based depressivity prediction models with different feature extraction and predictive methods.

The proposed methods were tested on the depressivity EEG dataset (introduced in Chapter 3). The EEG dataset includes eyes open and closed data epochs, and the data samples are separated into different gender groups. Similar to classification-based models, the performances of different feature extraction methods, regression methods, eyes states, and gender groups was compared in each set of models using statistical analysis (mainly ANOVA). All models' performance was also compared. These comparisons can give us: (i) a hint of how to choose suitable feature extraction methods and predictive analysis methods to predict depressivity using EEG data; (ii) a direction for designing EEG collecting experiments if we know which eyes state can produce better results; (iii) an idea of if there is a gender difference in depressivity EEG.

8.2 Overview

From the summary of related work (Sections 2.4.3.2 and 2.4.3.3), we can see that a lot of signal processing techniques are used for the feature extraction and various analysis methods to distinguish the features to classify the different kinds of brain-related diseases, neurological disorder diseases, and different physiological states. Therefore, the designs of Chapter 7 assume a distinction between health EEG and disease EEG and that is categorical. However, much more than classifying depressed patients and healthy controls, what we need is to predict the depressivity. To address the problem, this chapter introduces two regression-based depressivity prediction models to predict the depressivity based on EEG data. The regression models are based on a different view, which the EEG varies smoothly with a personality trait that is smoothly distributed in the population; with extreme scores representing 'sick' people. From this

point of view, there is no clear gap where you can put a dividing line between health and disease. Therefore, the key idea for model design in this chapter is that the EEG patterns vary progressively with progressive changes in depressivity.

Figure 3.3 in Chapter 3 shows that both kinds of models (regression-based and classification-based models) have the following establishment and evaluation steps: participant assignment, feature extraction, predictive analysis, depressivity estimation and model evaluation.

In the current regression-based models, k-fold cross-validation was used to choose training and testing data samples – as described in Section 3.3.1.2. Training data samples were used to train the prediction model, then the model was used to predict testing data samples' depressivity. Each epoch of testing data samples provided one prediction value. The ML-d score of each testing data sample was obtained by averaging its epochs' prediction values (Equation 3.3).

As with the classification-based models, R^2 = 11% was set as the threshold to decide if the model is likely useful for psychological application.

8.3 Regression Using FFT Features (Regression-Based Models I)

In this set of models, FFT was used as the feature extraction method which was exactly the same method as in models *Cla-FFTs* (Section 7.3.1). This set of models are *Reg-FFT-LSBoost*, *Reg-FFT-RF*, *Reg-FFT-Lasso*. They are collectively known as *Reg-FFTs*.

For model *Reg-FFT-LSBoost*, LSBoost was used as the predictive analysis method. The decision tree was used as the weak learner of LSBoost and the number of weak learners was 100. All decision trees were stumps.

For model *Reg-FFT-RF*, RF was used as the predictive analysis method. RF can be used to solve both classification and regression problems. In this model, it was used for regression. The decision tree was also used as the weak learner in this model. The factors which were use to control the depth of trees were set as the same in models *Cla-FFTs*. The number of features searched at each split point (m) was: $m = p/3$ (p is the number of all features). The details are provided in Section 6.3.1.3.

For model *Reg-FFT-Lasso*, lasso was used as the predictive analysis method. Lasso is an improved linear regression method. It has good prediction ability and can easily select important variables in the procedure of establishing a regression model. There-

fore, lasso is also used to establish a depressivity prediction model to check if it could provide higher prediction ability than other models. In this model, we used 5-fold cross-validation to determine the value of λ. The tolerance for the optimization was set to 0.0001 and the maximum iteration was 1000.

8.3.1 Experiments and Results

For different genders and eye states, nine experiments were implemented to predict participants' depressivity in each model (see Table 8.2). In Experiment (R1.1a-c), 44 female data samples were chosen. Each data sample includes approximately 230 eyes open epochs. The K-fold method was used in regression-based models, so 44 data samples were partitioned into 11 four-datasets sub-samples. In each round of the experiment, one sub-sample was left to test, and others were used to train the model. After 11 rounds, each epoch of each data sample was tested once. The ML-d score of each data sample was obtained by averaging all feature vectors' prediction values (Equation 3.3). These ML-d scores and their corresponding real PID-5 scores were then used to calculate the correlation value to evaluate the model. R1.2 - R1.9 (a-c) repeated the same procedures as Experiment (R1.1a-c), but used different data samples with different eye state epochs (shown in Table 8.3).

Gender	Eyes	*LSBoost*		*RF*		*LASSO*	
		Ex	R^2(*p-value*)	*Ex*	R^2(*p-value*)	*Ex*	R^2(*p-value*)
Female	Open	R1.1a	24.52% (0.0006)	R1.1b	11.70% (0.0230)	R1.1c	1.97% (0.3629)
	Closed	R1.2a	22.51% (0.0011)	R1.2b	10.84% (0.0291)	R1.2c	-0.01% (0.9452)
	All	R1.3a	**25.75%** (0.0004)	R1.3b	9.90% (0.0375)	R1.3c	0.33% (0.7122)
Male	Open	R1.4a	1.57% (0.5179)	R1.4b	**10.17%** (0.0917)	R1.4c	4.11% (0.2914)
	Closed	R1.5a	2.84% (0.3820)	R1.5b	0.99% (0.6069)	R1.5c	1.02% (0.6018)
	All	R1.6a	1.97% (0.4675)	R1.6b	8.36% (0.1282)	R1.6c	2.74% (0.3906)
Gender-Mixed	Open	R1.7a	**18.28%** (0.0002)	R1.7b	8.01% (0.0152)	R1.7c	-5.97% (0.0372)
	Closed	R1.8a	16.31% (0.0004)	R1.8b	10.32% (0.0056)	R1.8c	-29.06% (0.0000)
	All	R1.9a	14.37% (0.0009)	R1.9b	8.08% (0.0148)	R1.9c	-19.84% (0.0001)

Table 8.2: **Results of models *Reg-FFTs*.** The first and second columns of the table show the different genders and eye states for each experiment. Values in the table are R^2-values and the corresponding p-values (in brackets) of each experiment. The negative R^2-values represents the corresponding R-values were negative. *Ex* is the label of experiments.

Experiment	Gender (number of data samples)	Eye states (number of epochs/data sample)
R1.1 (a-c)	Female (44)	Open (~230)
R1.2 (a-c)	Female (44)	Closed (~230)
R1.3 (a-c)	Female (44)	both open and closed (~460)
R1.4 (a-c)	Male (29)	Open (~230)
R1.5 (a-c)	Male (29)	Closed (~230)
R1.6 (a-c)	Male (29)	both open and closed (~460)
R1.7 (a-c)	Both Female and Male (73)	open (~230)
R1.8 (a-c)	Both Female and Male (73)	closed (~230)
R1.9 (a-c)	Both Female and Male (73)	both open and closed (~460)

Table 8.3: Data samples and epochs used in each experiment in Table 8.2

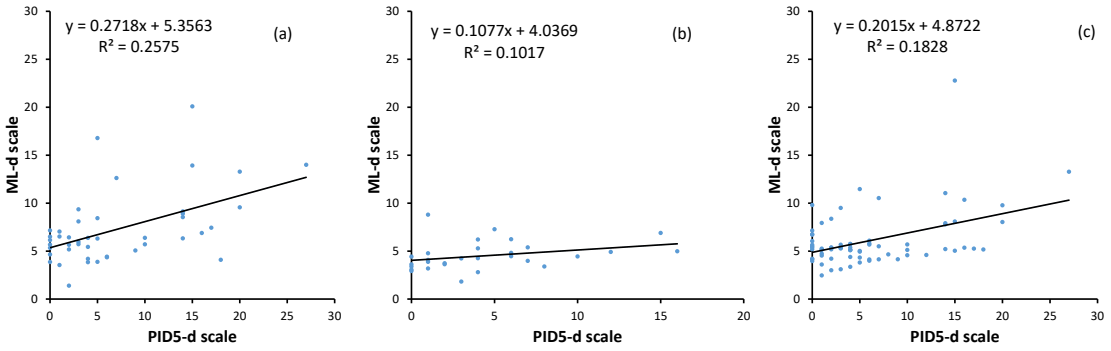

Figure 8.1: **Scatter plots of the best results of the first set of regression-based models.** Figure (a) shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.3a of model *Reg-FFT-LSBoost* which is the best result of female group; figure (b) shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.4b of model *Reg-FFT-RF* which is the best result of male group; figure (c) shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.7a of model *Reg-FFT-LSBoost* which is the best result of gender-mixed group. In figure (c), there is an apparent outlier (labeled by red circle), but delete it basically does not change the relationship (18.28% –> 18.51%).

For female group, the best result (25.75%) was obtained by using all eye states epochs with model *Reg-FFT-LSBoost*. This result is also the best one of all experiments. Figure 8.1a shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.3a of model *Reg-FFT-LSBoost*. For male group, the best result (10.17%) was obtained by using eyes open epochs of model *Reg-FFT-RF*. However, according to the p-value, the result is not significant. Figure 8.1b shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.4b of model *Reg-FFT-RF*. For gender-mixed group, the best result (18.28%) was obtained by using eyes open epochs of model *Reg-FFT-LSBoost*. Figure 8.1c shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.7a of model *Reg-FFT-LSBoost*. It is worth noting that the R^2-values obtained by model *Reg-FFT-Lasso* with gender-mixed group are significantly high negative values. Section 8.6.4 explained that this model fails to predict depressivity using EEG data because all ML-d scores are around 6. Modifying the parameters of lass does not improve the performance of this model. The high negative correlation was obtained by correlated PID5-d scores with the slightly fluctuated ML-d scores by chance.

In Table 8.2, there are three independent variables (factors). They are eye state, gender, and predictive analysis method. To test the effects of multiple factors in models, the R^2-values of all experiments were submitted to a three-way ANOVA using the highest order interaction as the error terms since there are no replicates. As can be seen in Table 8.2, the Lasso method is worse than other two predictive methods (Methods, $F(2, 8) = 43.79$, $p < 0.001$). Table 8.2 shows that different eye states had little influence on depressivity predictive ability (Eyes, $F(2,8) = 2.08$, $p = 0.187$). The results also show that for different genders, the performance of female group is much better than male and gender-mixed groups (Gender, $F(2,8) = 12.16$, $p = 0.0038$).

8.4 Regression Using CWT Features (Regression-Based Models II)

In this set of models, continuous wavelet transformation (CWT) was used to replace FFT as the feature extraction method. This set of models are *Reg-CWT-LSBoost*, *Reg-CWT-RF*, *Reg-CWT-Lasso*. They are collectively known as *Reg-CWTs*. The feature extraction methods used in this set of models were the same as the classification-based models *Cla-CWTs*. The predictive analysis methods used in this set of models were the same as the first set of regression-based models *Reg-FFTs*.

8.4.1 Experiments and Results

Because of the characteristics of CWT, the transferred data have two dimensions: frequency × time. In this model, we used two feature vector construction methods to transfer the two dimensions into one dimension. They were as follows: a) All values along the time domain were averaged to calculate accumulation of energy in each frequency slot; b) The top five maximum squared wavelet coefficient values along the time domain in each frequency slot were picked. The details were discussed in Section 5.3.4. Therefore, two groups experiments were implemented to predict participants' depressivity in this model.

Table 8.4 shows the R^2-values obtained by correlating the PID5-d scores against the ML-d scores of each participant in different experiments. The p-values were also calculated to see if the corresponding correlation was considered significant. R2.1 - R2.9 (a-c) used different data samples with different eye state epochs (As the same as the models *Reg-FFTs* which were shown in Table 8.3).

Compared to the first set of models (*Reg-FFTs*), the results of this set of models were worse even though the highest R^2-value was slightly better than the first set of models. For female group, the best result (27.55%) was obtained by using all epochs of model *Reg-CWT(Accumulation)-LSBoost*. This result is also the best one of all experiments and better than the one obtained of model *Reg-FFT-LSBoost*. Figure 8.2 shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R2a.3a of model *Reg-CWT(Accumulation)-LSBoost*. For male group, the best result (5.36%) was obtained by using all epochs of model *Reg-CWT(Accumulation)-RF*. But according to the p-value (0.2271), the result was not significant. For gender-mixed group, the best result (10.75%) was obtained by using eyes open epochs of model *Reg-CWT(Accumulation)-LSBoost*.

In the two parts of Table 8.4 ('Accumulation' and 'Top 5'), there are three independent variables (factors), respectively. They are eye state, gender, and predictive analysis method. To test the effects of multiple factors in models, the R^2-values of these two groups of experiments were submitted to a three-way ANOVA separately, using the highest order interaction as the error terms since there are no replicates.

As can be seen in the first part of Table 8.4 ('Accumulation' as feature vector construction method), the lasso method is much worse than other two predictive methods (Methods, $F(2, 8) = 21.03$, $p < 0.001$). The result also shows that the different eye states have little influence on depressivity prediction ability (Eyes, $F(2,8) = 2.6$, $p = 0.1345$). For different genders, the results obtained by male group are much worse

FVC	Gender	Eyes	*LSBoost*		*RF*		*LASSO*	
			Ex	R^2(*p-value*)	*Ex*	R^2(*p-value*)	*Ex*	R^2(*p-value*)
Accumulation	Female	Open	R2a.1a	15.20% (0.0089)	R2a.1b	13.41% (0.0145)	R2a.1c	2.50% (0.3058)
		Closed	R2a.2a	9.92% (0.0373)	R2a.2b	1.70% (0.3995)	R2a.2c	2.68% (0.2882)
		All	R2a.3a	**27.55%**(0.0003)	R2a.3b	6.43% (0.0967)	R2a.3c	2.61% (0.2952)
	Male	Open	R2a.4a	-0.78% (0.6490)	R2a.4b	3.94% (0.3018)	R2a.4c	-12.82% (0.0565)
		Closed	R2a.5a	0.65% (0.6781)	R2a.5b	-0.05% (0.9052)	R2a.5c	-16.93% (0.0266)
		All	R2a.6a	0.06% (0.9010)	R2a.6b	**5.36%** (0.2271)	R2a.6c	-14.05% (0.0451)
	Mixed	Open	R2a.7a	**10.75%**(0.0046)	R2a.7b	4.44% (0.0735)	R2a.7c	2.18% (0.2125)
		Closed	R2a.8a	4.91% (0.0595)	R2a.8b	1.42% (0.3145)	R2a.8c	2.42% (0.1890)
		All	R2a.9a	7.77% (0.0169)	R2a.9b	0.96% (0.4096)	R2a.9c	2.59% (0.1736)
Top 5	Female	Open	R2b.1a	8.45% (0.0557)	R2b.1b	10.31% (0.0336)	R2b.1c	2.14% (0.3433)
		Closed	R2b.2a	10.17%(0.0348)	R2b.2b	2.29% (0.3267)	R2b.2c	2.01% (0.3591)
		All	R2b.3a	18.56%(0.0035)	R2b.3b	4.70% (0.1575)	R2b.3c	2.23% (0.3331)
	Male	Open	R2b.4a	0.01% (0.9595)	R2b.4b	2.76% (0.3887)	R2b.4c	-10.65% (0.0841)
		close	R2b.5a	0.56% (0.6998)	R2b.5b	-0.16% (0.8374)	R2b.5c	-16.63% (0.0281)
		All	R2b.6a	0.06% (0.9015)	R2b.6b	2.59% (0.4040)	R2b.6c	-15.42% (0.0351)
	Mixed	Open	R2b.7a	0.77% (0.4606)	R2b.7b	2.31% (0.1994)	R2b.7c	2.15% (0.2161)
		Closed	R2b.8a	0.34% (0.6217)	R2b.8b	0.03% (0.8815)	R2b.8c	1.74% (0.2654)
		All	R2b.9a	1.57% (0.2910)	R2b.9b	0.07% (0.8200)	R2b.9c	2.15% (0.2159)

Table 8.4: **Results of models *Reg-CWTs*.** This table has two parts. The first half part lists the results which were obtained using 'Accumulation' feature vector construction method; the second part lists the results which were obtained using 'Top 5' feature vector construction method. Values in the table are R^2-values and the corresponding p-values (in brackets) of each experiment. The negative R^2-values represents the corresponding R-values were negative. *Ex* is the label of experiments.

than other two groups (Gender, F (4,8) = 25.88, p < 0.001).

The values of second part of Table 8.4 ('Top 5' as feature vector construction method) shows the similar information as the first part. The lasso method is much worse than other two predictive methods (Methods, F (2, 8) = 24.56, p < 0.001). The different eye states have little influence on depressivity predictive ability (Eyes, F (2,8) = 1.73, p = 0.2379) as well. For different genders, the results obtained by male group are much worse than other two groups (Gender, F (4,8) = 42.41, p < 0.001).

To compare these two feature vector construction methods, the values in Table 8.4

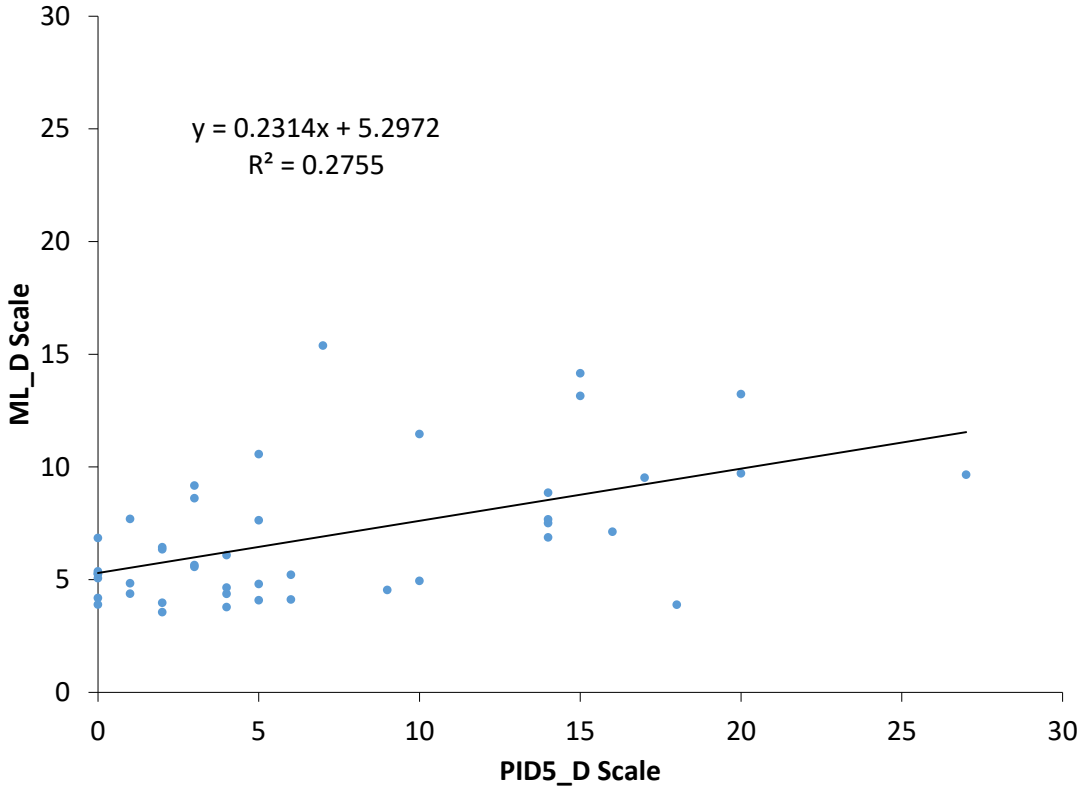

Figure 8.2: **Scatter plots of the best result of the second set of regression-based models.** This figure shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R2a.3a of model *Reg-CWT(Accumulation)-LSBoost* which is the best result of all experiments in this set of models.

were all submitted to a one-way ANOVA. As we can see from the table, the results of 'Top 5' are worse than 'Accumulation', but the difference is not huge (FVC, F (1,52) = 0.72, p= 0.3988).

8.5 Important Features

In the procedure of establishing depressivity prediction models, the important features which were most related to depressivity were discovered by LSBoost, RF, and lasso. In this section, the important features obtained by good models are shown.

Figure 8.3 shows the results of all regression-based models designed in this chapter except for the models using lasso as predictive analysis method. The results of models

using lasso are much worse than other two methods, therefore, the results were not compared in this chapter. We can see that, for most experiments with female and gender-mixed groups, model *Reg-FFT-LSBoost* produced the best results. For male group, most of the best results were obtained by model *Reg-FFT-RF*. Therefore, the important features obtained by these two models are described here.

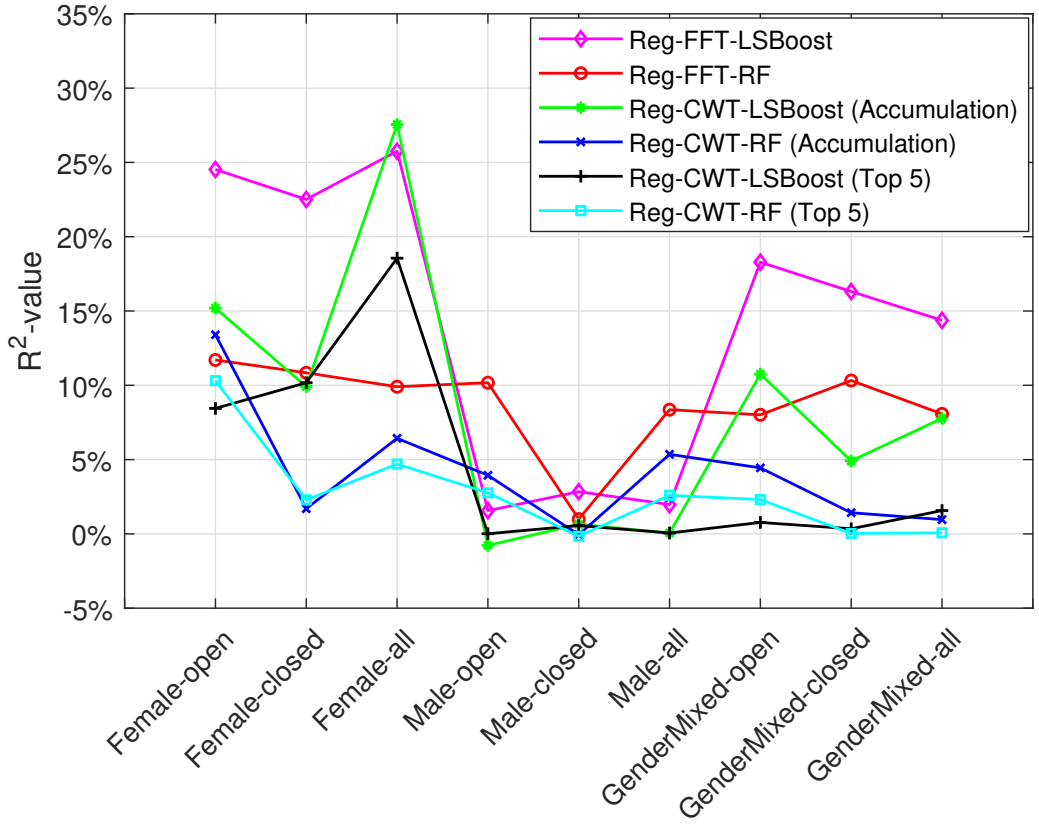

Figure 8.3: This figure shows the results of all regression-based models designed in this chapter.

Section 6.3.4 introduced the methods of calculating features' importances in LS-Boost and RF. After establishing an ensemble, a vector of weight values that indicates the relevant importance of each feature was obtained. The vector was then normalised between 0 to 1. The highest value in the vector indicates the most important feature. Because we implemented experiments with k-fold cross-validation, in each experiment, we established k ensembles. Therefore, we averaged the weight values across the k ensembles in the experiment to get the final rank of features' importance.

According to the components of FFT's feature vector (Figure 5.4), each point of

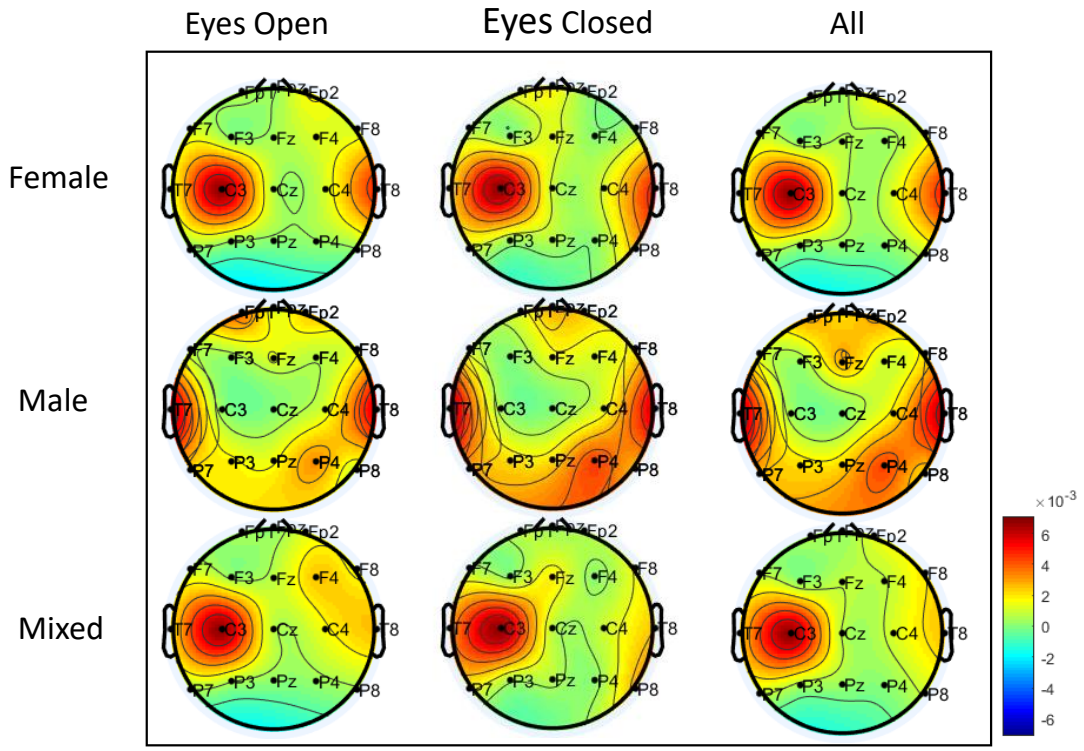

Figure 8.4: This figure shows the distribution of importance of different channels on the scalp. For female and gender-mixed groups, the results were obtained by model *Reg-FFT-LSBoost*; for male group, the results were obtained by model *Reg-FFT-RF*. From the figure, we can see that the most important channels of female group are C3, T7 and T8; the most important channels of male group are T7, T8, and frontal area; the most important channels of gender-mixed group are C3 and T7.

a feature vector contains two types of information: frequency and channel. The importance of different frequencies and channels were sorted by weight values separately. The distribution of important channels was shown in Figure 8.4, the figure shows that the most important channels of female group are C3, T7 and T8; the most important channels of male group are T7, T8, P4, and frontal area; the most important channels of gender-mixed group are C3 and T7. The distribution of important frequencies was shown in Figure 8.5. The figure shows that the most important frequencies are around 50Hz. For the female group, there are also peaks around 10Hz and 35Hz; for the male group, there are also peaks around 10Hz, and 70Hz; for gender-mixed with eyes open group, as well as 50Hz, the importance frequencies also distributed in 10Hz, 20Hz, 30Hz, and 40Hz. 30Hz, 35Hz, 40Hz, 50Hz and 70 hz are within the *gamma* band. The

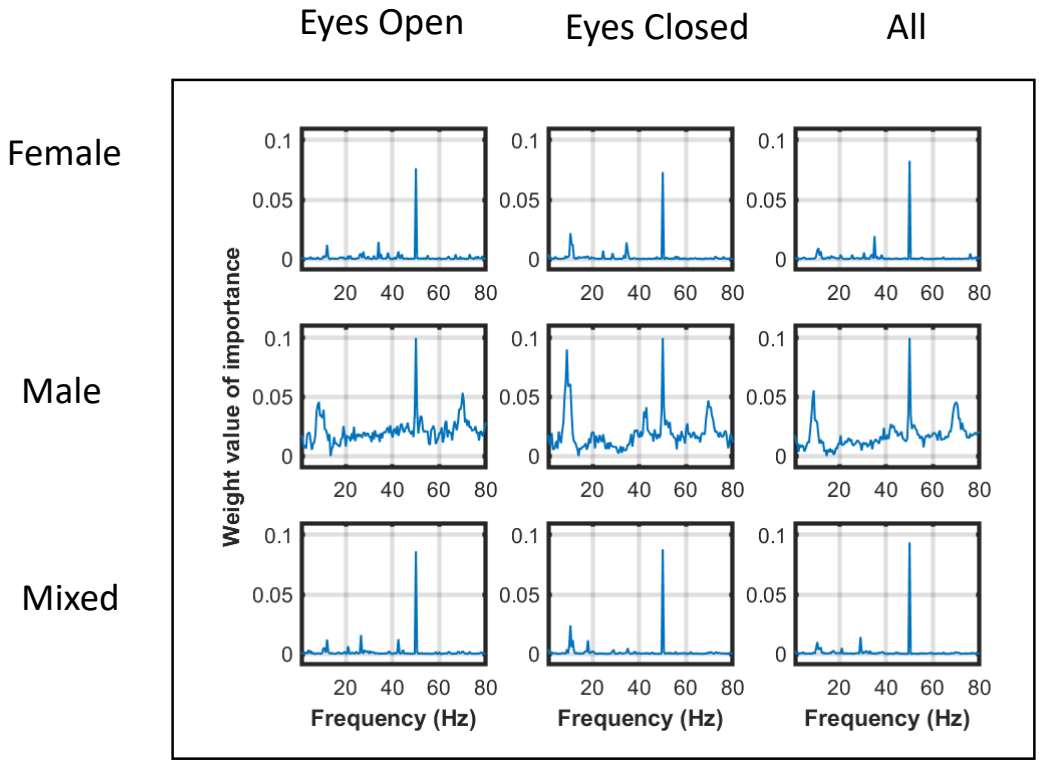

Figure 8.5: This figure shows the distribution of importance of different frequencies. For the female and gender-mixed groups, the results were obtained by model *Reg-FFT-LSBoost*; for the male group, the results were obtained by model *Reg-FFT-RF*. From the figure, we can see that the most important frequencies are around 50Hz. For the female group, there are also peaks around 10Hz and 35Hz; for the male group, there are also peaks around 10Hz, and 70Hz; for gender-mixed of eyes open group, except for 50Hz, the importance frequencies also distributed in 10Hz, 20Hz, 30Hz, and 40Hz.

gamma band is linked to memory, attention, and cognition. 10Hz is within the *alpha* band, this band is a frequently-used band for psychological study. 20Hz, which only occurs in the gender-mixed group, is belong to *beta* band. It should be noted that none of the R^2-values obtained by male group are significant, therefore, the features obtained by male group are not reliable.

Additional features with statistically significant R^2-values (p-value $<= 0.05$) obtained in the rest of the experiments are listed in Table 8.5. In the table, the important features marked with ** were obtained by experiments which obtained very reliable

R^2-values (p-value <= 0.001). We can see that for the female group, the optimal channels for predicting depressivity are C3, T7, and T8. For the gender-mixed group, the optimal channels for predicting depressivity are C3 and T7. The most important frequency band for these two groups is the *gamma* band. The *alpha* and *beta* bands also make some contributions. This set of channels and the frequency bands were all obtained from the model *Reg-FFT-LSBoost*. Another set of highly significant features was obtained from the model *Reg-CWT(Accumulation)-LSBoost* in the female group with all eye states epochs. The most important channels are, again, C3 and T8. The most important frequency band is the *gamma* band. The rest of the important features listed in Table 8.5 are almost all focused on T7, T8, and C3, and are in the *gamma* band. Some features also involve frontal areas (F4, Fp1) and P8 and C4, and the *alpha* band. Summing up the results, we find that C3, T7 and T8 are the most generally important channels and the *gamma* band is an important frequency band to predict depressivity. The *alpha* and *beta* bands may also make some contributions to predict depressivity.

However, we cannot confirm that the information obtained from F4, Fp1, P8 and C4 is of benefit to predict depressivity because we obtained these locations from the experiments with lower R^2-values. One possibility is that these locations bring interference to reduce the ability of depressivity prediction; another possibility is that these locations have limited contribution to predict depressivity, but these experiments failed to derive all the truly useful locations' information to cause the lower R^2-values. To figure this problem out, in future, we need to design more experiments such as feeding data without these channels into the models to check the change in R^2-values.

One important issue for interpretation is that we found frequencies around 50Hz are important for prediction of depressivity. We designed a phase-fixed mains noise removal method (see Section 3.4.2) to remove the mains noise and leave clean data within the same band. We also designed several experiments to verify that the results are based on real EEG signals and not noise. These experiments are discussed in the next section.

8.6 Validation

To verify and improve the results, a series of additional experiments were implemented.

Gender	Feature / Eyes / Model	Open		Closed		All	
		Channel	Freq (Hz)	Channel	Freq (Hz)	Channel	Freq (Hz)
Female	Reg-FFT-LSBoost	See Figure 8.4 & Figure 8.5**					
	Reg-FFT-RF	T8, T7	50, 12	T7, T8	50, 10.5	T7, T8	50, 12
	Reg-CWT(A)-LSBoost	C3, F4	68, 33	C3, Fp1	80, 48	C3, T8**	80, 64**
	Reg-CWT(A)-RF	T7, P8	79, 80	/	/	/	/
	Reg-CWT(T)-LSBoost	/	/	C3, Fp1	57, 79	C3, T8	57, 79
	Reg-CWT(T)-RF	T7, T8	79, 80	/	/	/	/
Mixed	Reg-FFT-LSBoost	See Figure 8.4 & Figure 8.5**					
	Reg-FFT-RF	T7, T8	50, 12	T7, T8	50, 10.5	T7, T8	50, 10.5
	Reg-CWT(A)-LSBoost	C3, C4	41, 80	/	/	C3, C4	80, 55
	Reg-CWT(A)-RF	/	/	/	/	/	/
	Reg-CWT(T)-LSBoost	/	/	/	/	/	/
	Reg-CWT(T)-RF	/	/	/	/	/	/

** p-value <=0.001

Table 8.5: Important features obtained from the regression-based models. All the features listed in this table were all obtained from the experiments with statistically significant R^2-values (p-value <= 0.05). The CWT(A) in the table means CWT(Accumulation); CWT(T) means CWT(Top 5); Freq (Hz) means Frequency (Hz).

8.6.1 About Mains Power

Figure 8.5 shows that 50Hz is one of the important frequencies to predict depressivity. However, the mains power in New Zealand is around 50Hz. To further verify the results, two additional experiments were implemented.

The first set of additional experiments — applied IIR notch filter: In the first set of additional experiments, an IIR (infinite impulse response) notch filter was applied to replace our designed phase-fixed mains noise removal method to filter the 50Hz and surrounding signals. Figure 8.6 shows the data process of this set of additional experiments. The left part of this figure shows the data preprocessing steps. Compared to the basic data preprocessing steps used in original experiments (shown in Figure 7.1), 50Hz notch filter is used to replace our designed phase-fixed mains noise removal method in this set of additional experiments. The lower cut-off frequency f1 of the notch filter is 49.751Hz and the upper cut-off frequency f2 is 50.251Hz. Figure 8.7 shows

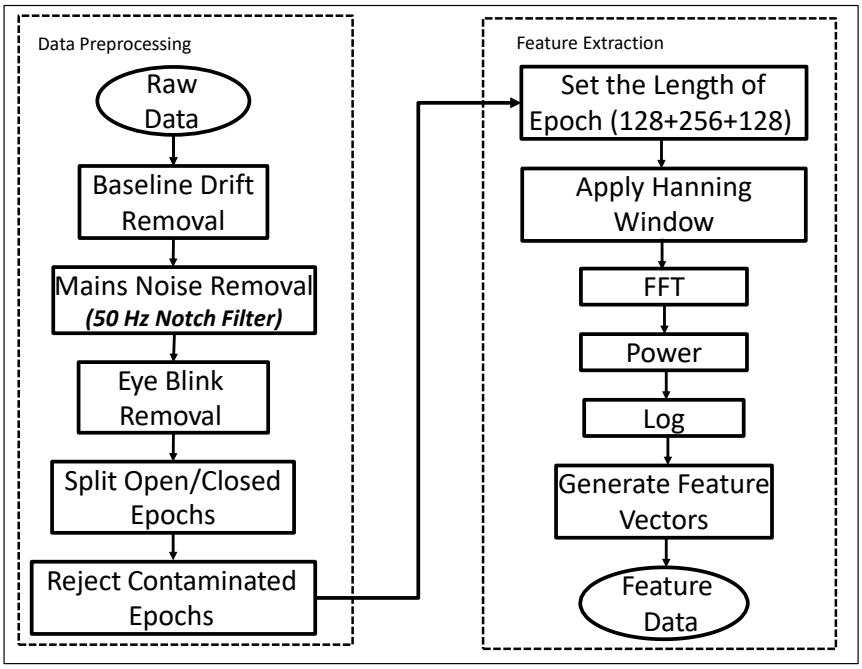

Figure 8.6: This figure shows the data process of the first set of additional experiments (notch filter). Compared to the basic data preprocessing steps used in original experiments (shown in Figure 7.1), the 50Hz notch filter replaced the mains noise removal method.

the notch filter located at 50Hz (0.39π radians/sample) with the desired bandwidth. Compared to phase-fixed mains noise removal method, the 50Hz notch filter creates band-holes while phase-fixed mains noise removal method only removes most of the 50Hz mains power.

Figure 8.3 shows the results of all regression-based models designed in this chapter. Model *Reg-FFT-LSBoost* produced the best results for most experiments of female and gender-mixed groups, and most of of model *Reg-FFT-RF* produced the best results for most experiments for male group. However, all results obtained of male groups are all not significant. Therefore, for female and gender-mixed groups, the remaining procedures of this set of experiments were the same as model *Reg-FFT-LSBoost*. The male group is not tested.

Table 8.6 shows the R^2-values of this set of experiments and the difference values which were obtained by subtracting the original R^2-values from the values of this set of experiments. This result shows that this set of experiments obtained lower R^2-values ($15.93\% \pm 5.83\%$) than original experiments ($20.29\% \pm 4.64\%$), especially the

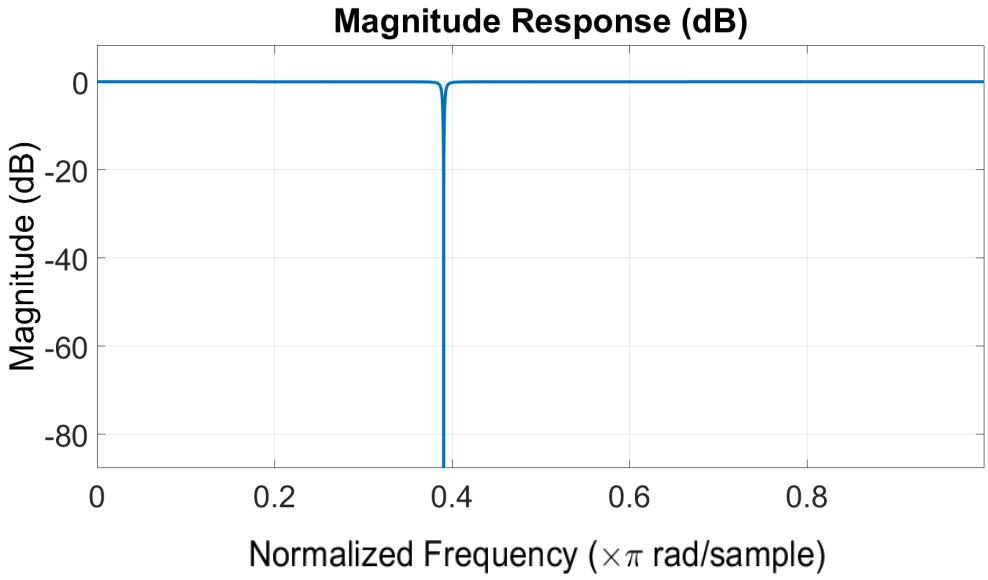

Figure 8.7: This figure shows that the designed notch filter was located at 50Hz (0.39π radians/sample) with the desired bandwidth.

Experiment A1	R^2-value	*Difference-value*
A1.1 Female - Eyes open epochs	23.09%	-1.43% (A1.1-R1.1a)
A1.2 Female - Eyes closed epochs	16.16%	-6.35% (A1.2-R1.2a)
A1.3 Female - All epochs	22.93%	-2.83% (A1.3-R1.3a)
A1.7 Gender-mixed - Eyes open epochs	11.19%	-7.09% (A1.7-R1.7a)
A1.8 Gender-mixed - Eyes closed epochs	10.62%	-5.69% (A1.8-R1.8a)
A1.9 Gender-mixed - All epochs	11.58%	-2.79% (A1.9-R1.9a)

Table 8.6: **Results of additional experiments — 50Hz notch filter.** R^2-values were calculated by correlating the PID-5 score against the ML-d scores of each participant in different experiments. The difference values were obtained by compared with the results of Experiment R1.1-1.3a, R1.4-1.6b, and R1.7-1.9a (Table 8.2).

R^2-values (11.13% ± 0.48%) of gender-mixed group were significantly lower than the original ones (16.32% ± 1.96%), t(4) = -4.4639, p = 0.0111.

The possible reason of the results is the notch filter creates band-holes around 50Hz, it might remove the effective information around 50Hz of the EEG signal. By contrast, the phase-fixed mains noise removal method that we designed only removes the mains

noise components, and the natural EEG was left.

The second set of additional experiments — interpolate 50Hz signal: The second set of additional experiments used a method called 'interpolate' to filter the 50Hz mains noise's influence. Figure 8.8 shows the data process of this set of additional experiments, and it illustrates the flow of raw data transferring to the feature data. The left part of this figure shows the data preprocessing steps. Compared to the basic data preprocessing steps used in original experiments (shown in Figure 7.1), the mains noise removal step was moved to the right part (feature extraction part). The right part of this figure shows the feature extraction steps, compared with the feature extraction steps of model *Reg-FFT-LSBoost* (shown in Figure 7.1), there was an extra step ('interpolate') in this set of additional experiments. It was used to interpolate 50Hz signal.

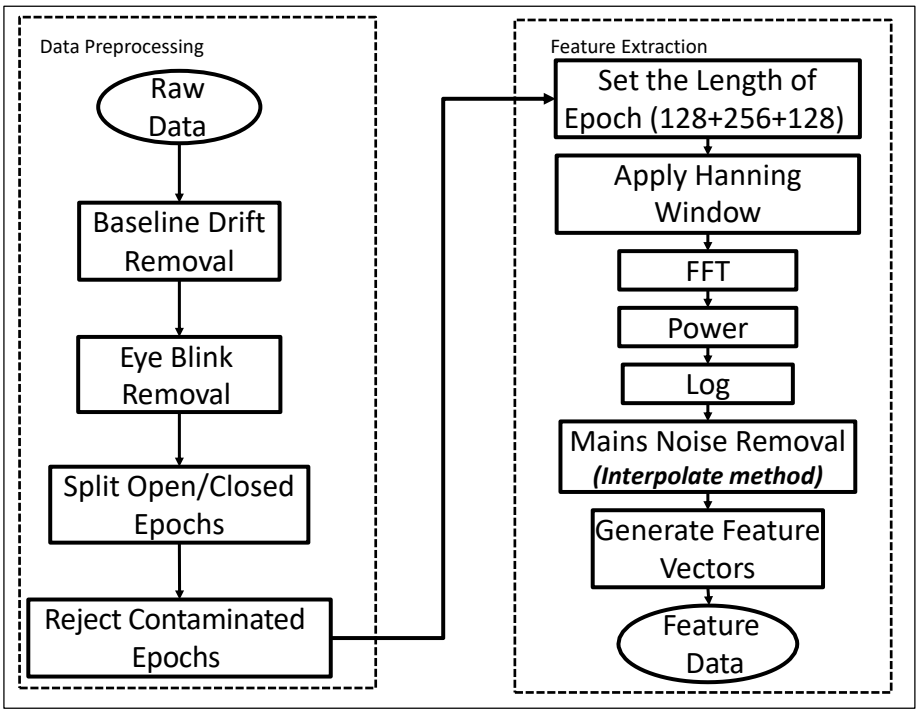

Figure 8.8: This figure shows the data process of the second set of additional experiments (interpolate). Compared with the original experiment 7.1, the mains noise removal step was discarded and the interpolate step was applied after the log transformation.

The 'interpolate' step was implemented by the following steps:

1. Replace the value at 50Hz with half of the sum of the value at 49Hz and the value at 51Hz. Shown in Equation 8.1.

$$Value\,(50Hz) = 1/2(Value\,(49Hz) + Value\,(51Hz)) \tag{8.1}$$

2. Replace the value at 49.5Hz with half of the sum of the value at 49Hz and the new value at 50Hz. Shown in Equation 8.2. The new value at 50Hz refers to the value calculated by Equation 8.1.

$$Value\,(49.5Hz) = 1/2(Value\,(49Hz) + New\,Value(50Hz)) \tag{8.2}$$

3. Replace the value at 50.5Hz with half of the sum of the new value at 50Hz and the value at 51Hz. Shown in Equation 8.3. The new value at 50Hz refers to the value that was calculated by Equation 8.1.

$$Value\,(50.5Hz) = 1/2(New\,Value\,(50Hz) + Value\,(51Hz)) \tag{8.3}$$

The remaining procedures of this set of experiments were the same as the first set of additional experiments mentioned previously.

Table 8.7 shows the R^2-values of this set of experiments and the difference value compared with the results of Experiment R1 (Table 8.2). This result also shows that this set of experiments obtained statistically significantly lower R^2-values (9.33% ± 5.57%) compared to original experiments (20.29% ± 4.64%), t(16) = -3.7033, p = 0.0041.

This experiment further shows that the effects at 50Hz are real. Because after tuning the values around 50Hz, it might remove the effective EEG signal.

Summary of additional experiments on mains power These two sets of experiments verified the results of the original experiments, providing evidence that the signals remaining after phase-fixed mains noise removal method were real EEG, rather than noise, and are important for predicting depressivity. In the first additional experiment, a notch filter was used to remove all 50Hz power (i.e. both noise and EEG) and, compared to phase-fixed removal, the R^2-values decreased. The notch filter creates a band-hole around 50 Hz, with a loss of any real EEG around 50Hz that we can conclude our phase-fixed mains noise removal method retained. In the second experiment, the interpolation method used the values at 49 Hz and 51Hz to replace the original vales at 49.5 Hz, 50 Hz and 50.5 Hz. Compared to the original results, the

Experiment A2	R^2-value	*Difference-value*
A2.1 Female - Eyes open epochs	16.24%	-8.28% (A2.1 - R1.1a)
A2.2 Female - Eyes closed epochs	9.67%	-12.84% (A2.2 - R1.2a)
A2.3 Female - All epochs	15.86%	-9.89% (A2.3 - R1.3a)
A2.7 Gender-mixed - Eyes open epochs	5.58%	-12.70% (A2.7 - R1.7a)
A2.8 Gender-mixed - Eyes closed epochs	4.76%	-11.55% (A2.8 - R1.8a)
A2.9 Gender-mixed - All epochs	3.88%	-10.49% (A2.9 - R1.9a)

Table 8.7: **Results of additional experiments — Interpolate.** R^2-value were calculated by correlating the PID-5 score against the ML-d scores of each participant in different experiments. The difference values were obtained by compared with the results of R1.1-1.3a, R1.4-1.6b, and R1.7-1.9a (Table 8.2).

R^2-values decreased sharply. The simplest explanation of this decrease is that the real EEG signal,containing the key information, around 50Hz was removed. In addition, our original R^2-values can be over 25%. The noise (either from mains power noise or introduced by mains noise removal method) should not be so highly correlated by PID5-d scores. These results show that the remaining signals after phase-fixed mains noise removal method were real EEG linked to depressivity.

8.6.2 Smooth

Another set of additional experiment was implemented to try to use traditional feature smoothing methods to improve the performance.

The third set of additional experiments — smooth In the third set of additional experiments, we tried to smooth the feature vector to improve the experiments' performance. Figure 8.9 shows the data process of this set of additional experiments, and it illustrates the flow of raw data transferring to feature data. The left part of this figure shows the data preprocessing steps. They were the same as the basic data preprocessing steps used in original experiments (shown in Figure 7.1). The right part of this figure shows the feature extraction steps, compared to the steps used in original experiments (shown in Figure 7.1), there is an extra step ('smooth') in this set of additional experiments. The remaining procedures of this set of experiments were the same as the first additional experiments mentioned previously.

The 'smooth' step was implemented by picking the median value of every three

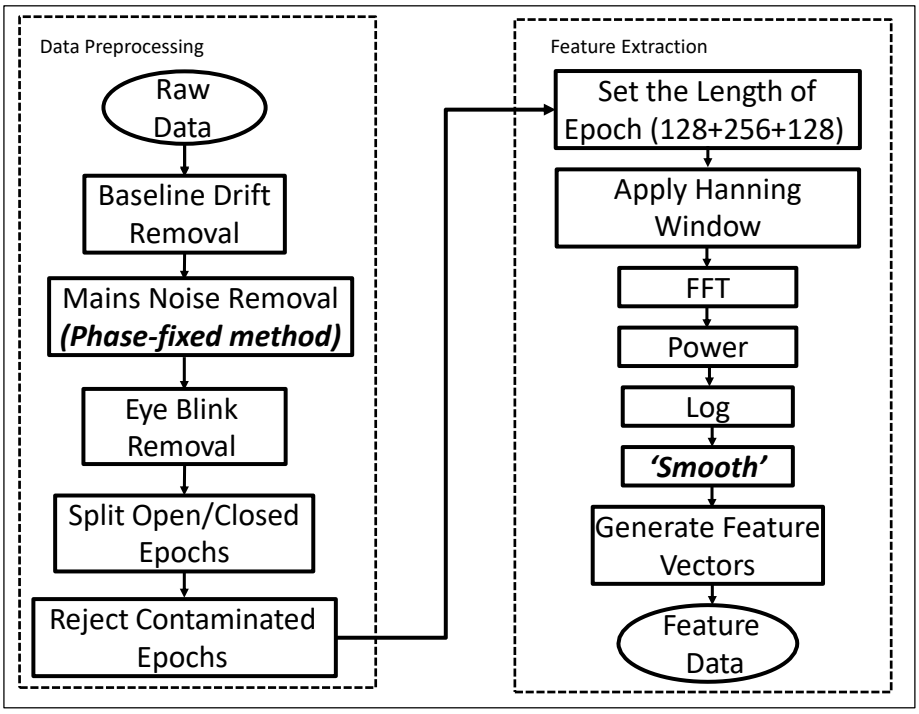

Figure 8.9: This figure shows the data process of the third set of additional experiments (smooth). The left part of this figure shows the data preprocessing steps. They were the same as the basic data preprocessing steps used in original experiments (shown in Figure 7.1). The right part of this figure shows the feature extraction steps, compared to the steps used in original experiments (shown in Figure 7.1), there is an extra step ('smooth') in this set of additional experiments.

original values. The blue solid line in Figure 8.10 shows a feature vector randomly picked from the original experiment, and it can be seen that there is jitter in the vector. The red dotted line in the figure shows the vector was treated after being smoothed; therefore, the line looks smooth.

Table 8.8 shows R^2-values of this set of experiments and the difference value compared with the original results. The result shows that most of the R^2-values (20.25% $\pm$ 5.27%) of this set of experiments were almost the same as original results (20.29% $\pm$ 4.64%), t(16) = -0.0153, p = 0.9881. The result implies that in this research, the 'smooth' method is not necessary.

This additional experiment confirmed that our designing models can choose features for itself and there is no need to do manual interventions.

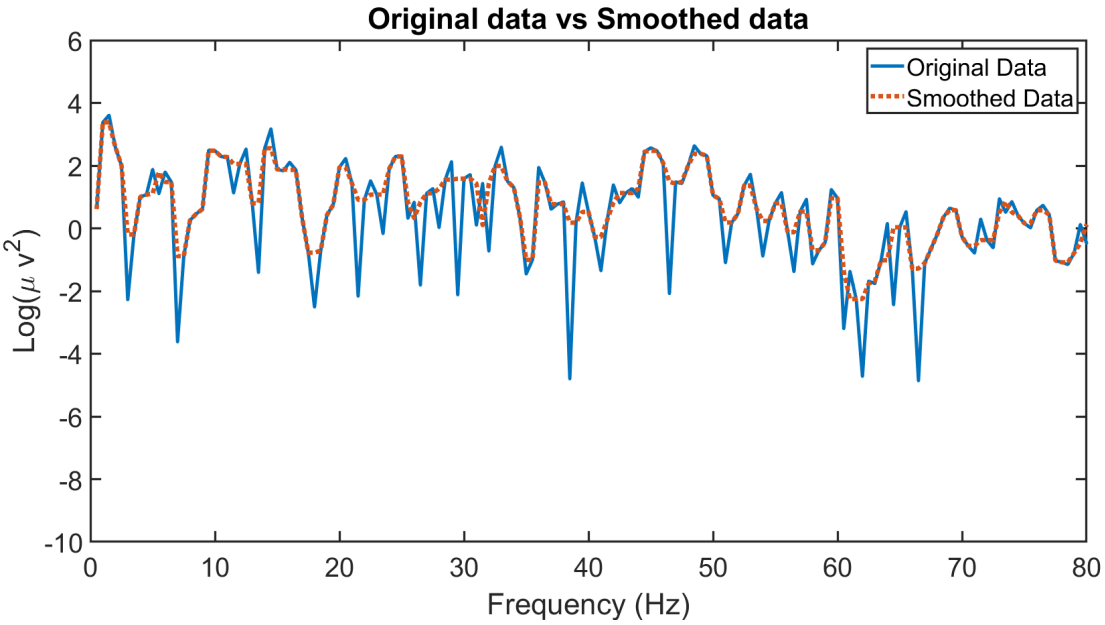

Figure 8.10: The blue solid line in this figure shows a feature vector randomly picked from the original experiment. It can be seen there is jitter in the vector. The red dotted line in the figure shows the vector was treated after being smoothed; therefore, the line looks smooth.

Experiment A3	R^2-value	*Difference-value*
A3.1 Female - Eyes open epochs	24.02%	-0.50% (A3.1 - R1.1a)
A3.2 Female - Eyes closed epochs	24.38%	1.87% (A3.2 - R1.2a)
A3.3 Female - All epochs	26.55%	0.80% (A3.3 - R1.3a)
A3.7 Gender-mixed - Eyes open epochs	14.87%	-3.42% (A3.7 - R1.7a)
A3.8 Gender-mixed - Eyes closed epochs	15.98%	-0.33% (A3.8 - R1.8a)
A3.9 Gender-mixed - All epochs	15.68%	1.31% (A3.9 - R1.9a)

Table 8.8: **Results of additional experiments — smooth.** R^2-values were calculated by correlating the PID-5 score against the ML-d scores of each participant in different experiments. The difference values were obtained by compared with the results of Experiment R1.1-1.3a, R1.4-1.6b, and R1.7-1.9a (Table 8.2).

8.6.3 Repeated-K-fold Cross-validation

In this chapter, we designed several regression-based models to predict depressivity. In the procedure of establishing models, the k-fold cross-validation method was used to select training and testing data samples. Cross-validation derives an accurate estimate

of model prediction performance. Therefore, it is a traditional approach to compare different models and select better ones (Seni and Elder, 2010; Jung, 2018). Schaffer (1993) confirmed that the cross-validation as a meta-learning technique could select the best prediction model from a set of models in most cases. However, cross-validation "gives a nearly unbiased but highly variable estimator" (Kim, 2009).

To avoid the variation, Bengio and Grandvalet (2004) pointed out that matched pairs cross-validation can be used to compare the performance of different algorithms, because matched pairs eliminates the additional variation caused by train/test splits. Therefore, in this thesis, we used matched pairs cross-validation to establish different regression-based models and compared their performance. Figure 8.3 shows the results of the regression-based models which achieved reasonable outputs in this chapter. We can see that, for female group, the best results were obtained by model *Reg-FFT-LSBoost*, the R^2-value are all over 22%; for male group, the best result was obtained by model *Reg-FFT-RF* using all eyes open epochs, the R^2-value is over 10%, however, the result is not significant (p-value> 0.05); for gender-mixed group, the best results were obtained by model *Reg-FFT-LSBoost*, the R^2-value are all over 14%.

However, from the application perspective, we need more robust R^2-values to measure the performance of proposed models and compared these results with the existing solutions. For this purpose we have selected the repeated-k-fold cross-validation method to test the proposed depression prediction models.

The four set of additional experiments — repeated-k-fold: The first step of the method is to shuffle the order of original data samples, and in the second step, applying the traditional k-fold cross-validation. The step one and two are repeated N times. As the same as our original test, after each k-fold cross-validation, we calculated the R^2-value between ML-d scores and PID5-d scores. After N times k-fold cross-validation, we obtained N R^2-values to evaluate models. Then the N values were averaged to obtain a robust R^2-value. In this research, we set $N = 100$.

As other additional experiments, the method was applied to the experiments which obtained the highest R^2-values of each group. For female and gender-mixed groups, the procedures of this set of experiments used model *Reg-FFT-LSBoost*; for male group, the procedures of this set of experiments used model *Reg-FFT-RF*.

Table 8.9 shows the average R^2-value with 95% confidence interval of 100 times repeated k-fold cross-validation. These results give us robust estimations of the models' performances to compare the results with other research. Compared with (Stewart and

Experiment A4	R^2-value	*Original Values*
A4.1 Female - Eyes open epochs	25.41%±1.41%	24.52% (R1.1a)
A4.2 Female - Eyes closed epochs	19.44%±1.32%	22.51%(R1.2a)
A4.3 Female - All epochs	23.84%±1.17%	25.75%(R1.3a)
A4.4 Male - Eyes open epochs	0.83%±0.77%	10.17% (R1.4b)
A4.5 Male - Eyes closed epochs	1.41%±0.68%	0.99% (R1.5b)
A4.6 Male - All epochs	0.82%±0.61%	8.36%(R1.6b)
A4.7 Gender-mixed - Eyes open epochs	16.78%±0.77%	18.28% (R1.7a)
A4.8 Gender-mixed - Eyes closed epochs	14.61%±0.68%	16.31% (R1.8a)
A4.9 Gender-mixed - All epochs	14.60%±0.54%	14.37%(R1.9a)

Table 8.9: **Results of additional experiments — repeated k-fold.** R^2-values were the mean±95% confidence intervals of 100 times repeated k-fold cross-validation's results. *Original Values* were the original R^2-values derived from Table 8.2.

Allen, 2018), the highest result we obtained is 25.41%±1.41% that is much better than 11%.

Compared the average values with original value (obtained by traditional k-fold cross-validation), we can see that for female groups, the repeated k-fold cross-validation obtained a little worse R^2-values (22.90% ± 3.09%) than the original experiments (24.26% ± 1.64%), t(4) = -0.6757, p = 0.5368 and the original values are in or close to the 95% confidence interval of repeated values; for gender-mixed groups the repeated k-fold cross-validation also obtained a little worse R^2-values (15.33% ± 1.26%) than the original experiments (16.32% ± 1.96%), t(4) = -0.7399, p = 0.5004 and the original values are in or close to the 95% confidence interval of repeated values; however, for male group, compared to the original value (6.51% ± 4.86%), the average R^2-values (1.02% ± 0.34%) of repeated k-fold cross-validation were decreased, t(4) = -1.9497, p = 0.123, and the original values are far away from the 95% confidence interval of repeated values. We can see that the results obtained from the female groups and the gender-mixed groups are more stable than the male group. It should be caused by the limitation of male group's size.

8.6.4 About Lasso

From Table 8.2 and 8.4, we can see that the results obtained by lasso are poor, especially for gender-mixed group of model *Reg-FFT-Lasso*, the R^2-value is -29.06%. Figure 8.11 shows the scatter plots of the ML-d score and the matched PID5-d score of the

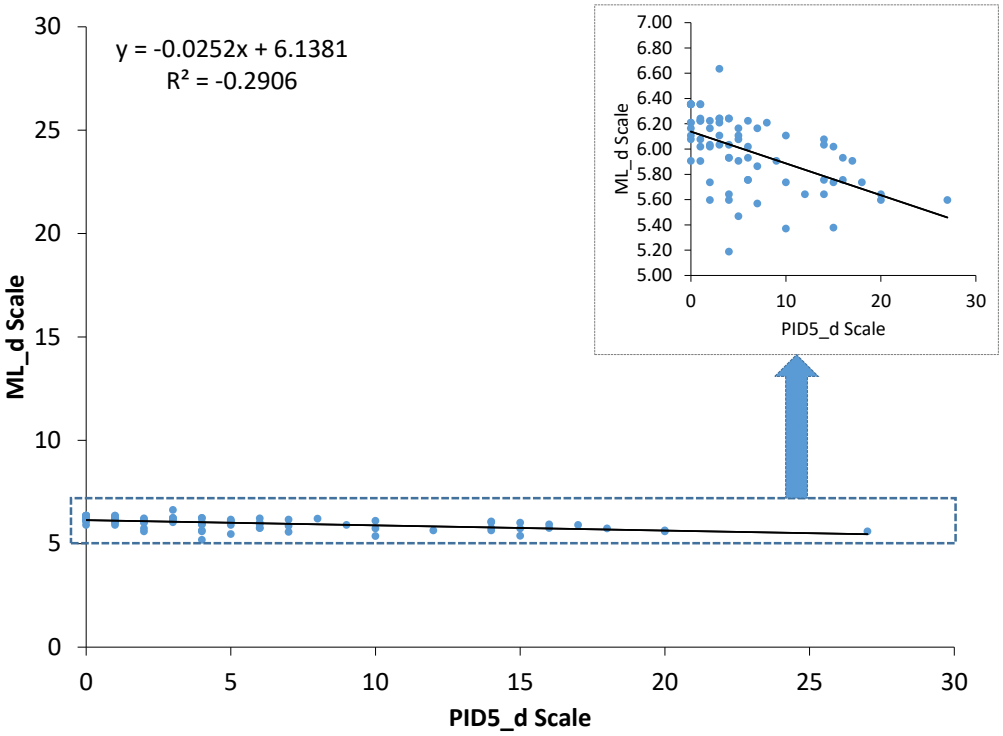

Figure 8.11: **Scatter plots of the worst result of model *Reg-FFT-Lasso*** This figure shows the scatter plots of the ML-d score and the matched PID5-d score of the experiment R1.8c of model *Reg-FFT-Lasso* which is the worst result of all experiments in this model.

experiment R1.8c of model *Reg-FFT-Lasso* which is the worst result of all experiments in this model. From the partial enlarged detail of Figure 8.11, we can see that all ML-d scores are in the range of 5 to 7. It means that lasso failed to establish depressivity prediction model, and the model is arbitrarily worse. The high negative correlation was obtained by correlated PID5-d scores with the slightly fluctuated ML-d scores by chance. In such a case, the R^2-value does not reflect the relationship between ML-d scores and PID5-d scores.

To validate the performance of lasso, some adjustments were made to see if they would improve the lasso's performance. In the original experiment, the value of λ

was selected by 5-fold cross-validation. The error tolerance was set to 0.0001 and the maximum iteration was 1000. In some rounds, the least square objective didn't converge to the error tolerance before the maximum iteration was reached. To improve the performance of model, we increased the maximum iteration to 1e5 and rerun all the experiments. However, this set of experiments produced almost the same results as original experiments. It means that in this case, 1000 iteration is enough. To increase iteration does not help to improve the model.

Experiment A5	R^2-value	*Difference-value*
A5.1 Female - Eyes open epochs	2.10%	0.13% (A5.1-R1.1c)
A5.2 Female - Eyes closed epochs	-0.04%	-0.03% (A5.2-R1.2c)
A5.3 Female - All epochs	0.36%	0.03% (A5.3-R1.3c)
A5.4 Male - Eyes open epochs	3.99%	-0.12% (A5.4-R1.4c)
A5.5 Male - Eyes closed epochs	0.93%	-0.09% (A5.5-R1.5c)
A5.6 Male - All epochs	2.41%	-0.33% (A5.6-R1.6c)
A5.7 Gender-mixed - Eyes open epochs	-5.97%	0.00% (A5.7-R1.7c)
A5.8 Gender-mixed - Eyes closed epochs	-29.18%	-0.12% (A5.8-R1.8c)
A5.9 Gender-mixed - All epochs	-20.44%	-0.60% (A5.9-R1.9c)

Table 8.10: **Results of additional experiments — about lasso.** R^2-values were calculated by correlating the PID-5 score against the ML-d scores of each participant in different experiments. The difference values were obtained by compared with the results of Experiment R1.9c (Table 8.2).

Because high λ value can cause underfitting, and low λ can cause overfitting. It is important to choose proper λ. In original experiment, we used 5-fold cross-validation to choose λ in the range of $(0.001 * \lambda_{max}, \lambda_{max})$. The λ_{max} was defined by Equation 6.18. We extended the searching range of λ to see if it will improve the model's performance. The λ_{min} was set as $1e-5 * \lambda_{max}$. Then the searching range was equally divided into 1000 points on the logarithmic scale. Table 8.10 shows R^2-values of this set of experiments and the difference value compared with the original results. We can see the results of this set of experiments were still not improved.

These results suggested that lasso as a linear regression method does not fit to work with the features extracted from depressivity EEG using FFT and CWT methods. It implies that the depressivity EEG feature data have a non-linear shape. The better results obtained by ensemble methods based on decision trees also demonstrates the

non-linear shape of EEG feature data.

8.7 Conclusion

In this chapter, two sets of regression-based models were designed to predict depressivity using EEG data and a series of experiments were implemented to test the performance of these models.

Figure 8.3 shows that the best results of these experiments were almost all obtained by model *Reg-FFTs*. We implemented 100 times repeated-k-fold cross-validation to get solid R^2-values. The results listed in Table 8.9 show that the best result is 25.41%±1.41% obtained in the female group with eyes open epochs, which doubles the result (11%) reported by previous research (Stewart and Allen, 2018).

The weight values calculated by LSBoost and RF algorithm show that the most important channels for the female group are C3, T7 and T8; the most important channels for the male group are T7, T8, P4, and the frontal area; the most important channels for the gender-mixed group are C3 and T7. The important frequencies are mostly in the *gamma* band (30, 35, 40, 50, 70Hz). The results also showed that 10Hz and 20Hz contributed to depressivity prediction. The summary of all important features listed in Table 8.5 imply that C3, T7 and T8 are the most generally important channels and the *gamma* band is an important frequency band to predict depressivity. The *alpha* and *beta* bands may also make some contributions to predict depressivity. This conclusion is useful for psychologists to further explore the personality (like depressivity, anxiety) focus on these locations and frequencies. It should be noted that the R^2-values of the male group are very low, the important locations and frequencies of the male group are not very reliable.

From the results, we can see that the most important frequencies include 50Hz. We designed a series of additional experiments, which removed the mains noise component and showed that the critical 50Hz components for depressivity are real brain waves rather than mains noise. Meanwhile, the additional experiments showed that our design methods can choose important features without manual feature processes.

Other outcomes and with other related work are discussed in the last chapter (Chapter 9).

Chapter 9

Conclusions and Future Work

9.1 Conclusions of This Thesis

EEG measures electrical signals at the scalp and can give information about brain processes near the surface of the brain (cortex). It is potentially an important tool in the diagnosis of emotional states, brain-related diseases and other physiological states. However, there are massive amounts of data in the EEG and it is hard to analyse these data visually. Therefore, using machine learning methods to develop automatic EEG analysis methods is important. In this thesis we focused on the relation of depressivity to EEG.

Previous researchers have already designed a number of methods to distinguish clinical depressed patients from normal ones via EEG. However, simply distinguishing clinical patients from normal people is not enough, because the effects of mental treatments depend on the treatment's starting time. If the treatment starts early, it will be more effective (Hosseinifard *et al.*, 2013; Mohammadi *et al.*, 2015). Therefore, to predict the depressivity of people in general rather than just classifying depressed patients is important.

In this thesis, we first developed some methods to improve the EEG signal quality. Then, we designed several models to predict depressivity via EEG signals to help prescreening for depression. In other words, the model could be considered as measuring the personality of the type person who has the potential to become depressed. The following are three main objectives that this thesis focussed on:

(1) To develop automatic models to predict depressivity using EEG data and investigate which model is suitable for predicting;

(2) To investigate which EEG data (EEG channel and frequency) are suitable for predicting depressivity;

(3) To develop methods to improve EEG artefact removal, especially for removing the eye blink and mains noise.

9.1.1 First Objective: Develop Depressivity Prediction Models

In recent years, EEG studies on depression have shown that EEG data can be used to effectively distinguish depressive participants from healthy controls. Some research has reported that the classification accuracies of classifying depressive participants from healthy controls using machine learning technologies have already been over 90% (summarised in Table 2.2). However, as mentioned above, only distinguishing clinical depressed patient from normal ones is not enough.

Therefore, the first objective of this research was to develop automatic prediction models to predict depressivity. To achieve this objective, we proposed two main types of models: classification-based models, based on the assumption that the appearance of some specific EEG patterns which indicate High depressivity increases in number with a higher depressivity; and regression-based models, based on the assumption that EEG patterns vary with different depressivity. The best result obtained by regression-based models was 25.41%±1.41% (mean±95% confidence intervals of 100 times repeated k-fold cross validation's results). There are very few researchers doing similar work to us in testing personality using EEG data, with previous work focussing almost entirely on classifying clinically depressed people from normal (see Chapter 2). In our lab, the same dataset used in this research with an additional 28 participants was tested using traditional EEG alpha asymmetry and HFD (Higuchi's fractal dimension) with stepwise regression. The correlations obtained by the traditional EEG alpha asymmetry method were very low and showed no significant correlation between alpha asymmetry and depressivity. The best R^2-value was obtained by HFD and was 4%. In a previous paper, Stewart and Allen (2018) reported a longitudinal pilot study where they examined the relationship between resting frontal EEG asymmetry and BDI scores (BDI is a clinical scale for depression), the highest R^2-values obtained by their female group was only 11%. It looks like our methods are better than the traditional ones. Furthermore, our results are based on the PID5-d scale which covers the normal range (maps to personality, not just sickness), therefore, our method is more suitable for

detecting depressivity as a trait than the previously reported one.

We used different feature extraction and predictive methods to establish the models. In the classification-based models, we used FFT, CWT, CSP, and FFT Combining CSP to extract features from the EEG data. AdaBoost, RF, and LVQ were used separately as predictive analysis methods to analyse features and establish models. Because CSP, AdaBoost, and LVQ were only suitable for the classification problem, they were not used in the regression-based models. In the regression-based models, only FFT and CWT were used to extract features. LSBoost, as another kind of boosting method, was used with RF and Lasso to establish regression-based models.

9.1.1.1 Comparison of design assumptions

The best result from the classification-based models with the gender-mixed group was 17.80% (Cla-CSP-RF), while the best result from the regression-based models was 16.78%±0.77% (Reg-FFT-LSBoost). These two results are close.

CSP cannot be applied to regression models. So, for a fair comparison, the result obtained from the model *Cla-FFT-RF* (10.95%) was compared with the result obtained from the model *Reg-FFT-LSBoost* (16.78%±0.77%). Both were obtained from the gender-mixed group. The comparison suggests that the performance of regression-based models is better than the classification-based models.

It is worth noting that from the regression-based models, the pure female group obtained better result (25.41%±1.41%) than gender-mixed group. Whereas in the classification-based models, data samples were not split into different gender groups because of the limitation in the size of the dataset. In future, we need to do more experiments using more data.

9.1.1.2 Comparison of feature extraction methods

FFT and CWT as two time-frequency methods were used to extract frequency domain features from EEG for depressivity prediction.

For both sets of models, the overall results with FFT are better than CWT (Accumulation), and the results with CWT (Accumulation) are better than CWT (Top 5). CWT (Accumulation) is similar to FFT, because CWT (Accumulation) calculates the accumulation of energy in each frequency slot. But they used different functions to transfer the signal into the frequency domain and have different frequency resolutions. The results show that the frequency components obtained by FFT are more useful to reflect the intrinsic information of the signal.

CWT (Top 5) selects the five largest wavelet coefficient values in each frequency slot, which reflect the five top sudden changes of EEG signal in each frequency of each epoch. This method extracts the singularities of the EEG signal in each epoch to find if different depressivity EEG signals have different sudden transients. However, the results of CWT (Top 5) are much worse than FFT and CWT (Accumulation). The poor behaviour of CWT (Top 5) demonstrates that depressivity is a personality state which generates EEG components that do not change in a short time. Therefore, it is harder to characterize different depressivity EEG signals by different sudden transients than by ongoing rhythms.

In the classification-based models, CSP was also adopted to extract information from the EEG signal because of its good spatial distinguishability. Comparing with other signal processing methods used in the classification-based models, the method of CSP produced the highest R^2-value (17.80%). The highest result obtained by two time-frequency methods in the classification-based models was only 10.95%. The result obtained by CSP was much better than using FFT and CWT. The result demonstrated that CSP is good at extracting information from different EEG conditions. Furthermore, the FFT combined with CSP obtained worse results than pure CSP. This implies that the spatial distribution information obtained by the time domain signal is a better predictor of depressivity than the spatial distribution information obtained by the frequency domain signal.

The overall results obtained by the classification model implied that the features obtained by the frequency domain were worse than the features obtained by the original time domain signal. Since CSP cannot be applied to the regression models, the current results of the regression models were all based on the features of the frequency domain. In future, we will try to apply other feature extraction methods which are comparable to CSP (e.g. PCA) to the regression-based models. Since we have concluded that the performance of the regression model is better than the classification model in Section 9.1.1.1, the results with new feature extraction methods are expected to be better.

9.1.1.3 Comparison of predictive methods

In the classification-based models, the performance of the two ensemble methods is similar (RF is slightly better than AdaBoost), but all are much better than LVQ (a neuron network). This could be because that the ensemble methods combine several weak learners as strong learners, the ensembles have a strong ability to improve classification (Breiman, 2001). LVQ needs to set appropriate parameters, like the number of

neurons, initialization of the weight matrix (Spiliopoulou, Kruse, Borgelt, Nürnberger, and Gaul, 2006). The classification accuracy of LVQ is related to the features of data set. This sometimes makes the performance of LVQ unstable.

In the regression-based models, the performance of the two ensemble methods is much better than lasso. This shows again that the ensemble methods are good methods to predict depressivity. The poor results of lasso demonstrate that lasso as a linear regression method does not work with the features extracted from depressivity EEG using FFT and CWT methods. It implies that the depressivity EEG feature data have a non-linear shape. The better results obtained by ensemble methods based on decision trees also demonstrates the non-linear shape of EEG feature data.

Therefore, in this research, ensemble learning as a powerful tool for data analysis was demonstrated to be more suitable than the traditional neural network or linear regression to predict depressivity.

Comparing two ensemble methods, in the regression-based model, in most cases, the performance of LSBoost is better than RF while in the classification-based model, RF is slightly better than AdaBoost in most cases. The results are consistent with the conclusion of Breiman (2001) that in classification, the performance of RF is as good as AdaBoost, and sometimes RF is better; while in regression, the RF produces less good results than in classification. This could explain why in the regression-based model, the performances of LSBoost is better than RF.

9.1.1.4 Comparison of genders

There are many reports of gender differences in the causes of depression, and gender differences in the EEG of depressed people. Halbreich and Lumley (1993) suggested that the differences in mechanisms of depressive symptoms between male and female might be related to the central nervous system. Bryden (1982) reported that gender may modulate hemispheric EEG asymmetry, while Smit *et al.* (2007) suggested that EEG asymmetry may only relate to depression in young females rather than males. Ahmadlou, Adeli, and Adeli (2013) found significant differences between male and female adults with major depressive disorder in relative convergence of *delta* band EEG in the intraleft temporal and frontoleft temporal lobe. Males and females with depression also appear to have different slow-wave activities during non-rapid eye movement sleep (Reynolds III, Kupfer, Thase, Frank, Jarrett, Coble, Hoch, Buysse, Simons, and Houck, 1990; Armitage, Hoffmann, Trivedi, and Rush, 2000; Armitage and Hoffmann, 2001). To demonstrate the gender difference, in regression-based models, the dataset

was split by gender to implement experiments.

The performance of the models in the male group is very poor. However, the main possible reasons are we only have 29 data sets in the male group and the PID5-d range of the male group in our dataset was narrow. Therefore, it cannot easily demonstrate a gender difference or lack thereof.

The results show that the performance of the gender-mixed group is worse than the female group. To an extent, this result confirmed the existence of a gender difference. The different EEG patterns from different genders lower the model's depressivity prediction ability at least for regression models.

9.1.1.5 Comparison of eye states

Barry, Clarke, Johnstone, Magee, and Rushby (2007) demonstrated that the eyes closed and eyes open states provide EEG measures differing in topography and power levels. When the eyes are closed, cortical activation is decreased (Gilbert, McClernon, Rabinovich, Sugai, Plath, Asgaard, Zuo, Huggenvik, and Botros, 2004). Because the alpha power is usually more dominant when the eyes are closed (Lindgren, Larson, Schaefer, Abercrombie, Ward, Oakes, Holden, Perlman, Benca, and Davidson, 1999; Bruder *et al.*, 1997) and the previous depression study usually focussed on the alpha band, some research only used the eyes closed data (Baehr *et al.*, 1998; Knott *et al.*, 2001; Kemp, Griffiths, Felmingham, Shankman, Drinkenburg, Arns, Clark, and Bryant, 2010).

However, most results obtained from both classification- and regression-based models show that compared to the eyes closed state, depressivity was more sensitive to the EEG data collected with eyes open state. This may provide a benefit for designing EEG collecting experiments because participants don't need to be constrained to close their eyes. However, the two methods would need to be directly compared in case eye closure changes the eyes open EEG during relaxation testing.

In most cases, the results of using all eye states epochs are better than the eyes closed epochs and worse than the eyes open epochs. This is reasonable because the bad performance of the eyes closed epochs should lower the performance of eyes open epochs. However, in some cases (e.g. model *Cla-CSPs*), the performances of all eye states are better than both eyes closed state and eyes open state. The most likely reason is that the algorithm extracted common features from both eye state epochs and the number of all eye states training epochs is double the number of single eye state (closed or open) training epochs. Therefore, in these cases, the R^2-values obtained by all eye states epochs are higher than the R^2-values obtained by eyes closed epochs or

eyes open epochs.

9.1.2 Second Objective: Investigate Which EEG Data (EEG Channel and Frequency) Are Suitable for Predicting Depressivity

In this study, we used ensemble learning as our main predictive method. The most important advantage of ensemble learning is its easy interpretation, therefore, we could easily find the most important features that the ensemble chose. In classification models, we introduced CSP as the feature extraction method to find the spatial patterns of EEG which can distinguish two kinds of EEG data. From the patterns obtained by CSP, we could find locations of the critical features to distinguish two kinds of data. These features are the potential biomarkers we tried to find.

9.1.2.1 Channels

EEG research on depression focuses primarily on the frontal field (e.g.,Henriques and Davidson, 1991; Allen *et al.*, 2004; Segrave *et al.*, 2011; Ahmadlou *et al.*, 2012a). Our result also showed that the EEG of the frontal field has a close relationship with depressivity. The common spatial pattern (including Fp2) obtained in the classification-based model and the extraction of the frontal areas by the regression-based model of the male group showed that the EEG of the frontal field has a close relationship with depressivity. But it should be noted that the R^2-values of the male group are not high, the important locations and frequencies of the male group are not very reliable.

Furthermore, the result obtained from the regression-base model of female group are C3, T7 and T8; the most important channels of gender-mixed group are C3 and T7. The common spatial pattern also included T7. These results, taken together, show that the EEG of channel T7 and T8 are important to predict depressivity. This result is consistent with recent research by Liao, Wu, Huang, Cheng, and Liu (2017). Their best EEG classification accuracy for classifying depressed patients was obtained by 8-electrode EEGs of the temporal area. Compared with that research, our outcome narrowed the range from 8 electrodes to 2 electrodes.

Our research also found another specific location – C3 – which is important for depressivity. This location has not been previously reported to have a relationship with depressivity and so is a new finding. Given its location, it may be linked to the tendency for T7 to be detected more than T8.

9.1.2.2 Frequency

In previous EEG research on depression, the *alpha* band (8–13Hz) has been most used (e.g.,Baehr *et al.*, 1998; Kemp *et al.*, 2010; Segrave *et al.*, 2011; Gordon *et al.*, 2010). Consistent with this, we found that signals in the region of 10Hz, which is in the *alpha* band, contributed to depressivity prediction. There is little research on the relationship between the *beta* band and depression. Merkl, Neumann, Huebl, Aust, Horn, Krauss, Dziobek, Kuhn, Schneider, Bajbouj, *et al.* (2015) found difference in the *beta* band between depressed patients' reported emotional empathy for negative stimuli and patients reported to have no empathy. Sheikhani, Behnam, Mohammadi, Noroozian, and Golabi (2008) reported that the signal obtained from the *beta* band can be used to distinguish Autism disorders from controls. We found that signals around 20Hz, which is in the *beta* band, are also useful to predict depressivity.

Previous study indicated that the *gamma* band is suitable for EEG-based emotion classification (Li and Lu, 2009). Webb, Dillon, Pechtel, Goer, Murray, Huys, Fava, McGrath, Weissman, Parsey, *et al.* (2016) found that the EEG of the *gamma* band were highly associated with three promising endophenotypes of depression. The results from Liao *et al.* (2017) also showed that the EEG of *gamma* bands is a good indicator for classifying depressed patients and controls. From the results of our research, the *gamma* band (around 30Hz, 35Hz, 40Hz, 50Hz, 70Hz) was found to be an important band for measuring depressivity. This result is consistent with previous research.

9.1.3 Third Objective: EEG Artefact Removal

Human EEG contains many complex components and is always contaminated by artefacts, with particularly large components arising from eye blinks. Recent research (summarised in Section 2.4.3.1) has reported many largely effective artefact removal methods. However, some of the artefact removal methods for eye blinks need extra electrodes to record the eyes' activities, and this will increase hardware cost. Some other methods, such as ICA, which do not need extra electrodes, normally cannot handle single channel signals. In this thesis, we developed an automatic eye blink artefact removal method based on a novel, physiologically-inspired, template approach.

The method we proposed can detect and remove eye blink components based on analysis of a single channel and with a success rate of over 90% in terms of recovering the variance of the original EEG prior to the addition of the eye blink components.

Mains power noise is another artefact that EEG signal can always be contami-

nated from electrical equipments. Section 3.4.2 describes a method that we designed to remove the fixed-phase mains power noise and leave a clean signal with residual 50Hz EEG. The result shows that after using this method, the mains power noise was removed.

9.2 Future work

In future, we will explore the possibility of extending the fundamentally template-based eye blink removal, and mains power noise removal method to other EEG artefacts. The eye blink artefacts removal method has potential to be used online and it can be implemented on each channel independently. Therefore, it should be particularly suited for use in portable brain-computer-interfaces and in neurofeedback training. The important brain location and signal frequency for measuring depressivity we found are also useful to design portable device because we only need to focus on a small number of channels and one frequency band. The finding that the eyes open state's EEG is more sensitive to depressivity is already being used in neurofeedback training in our lab. Therefore, in future, the combination of the findings and the methods designed in this thesis can be used to design portable devices to help estimate depressivity and provide neurofeedback training.

We believe the models designed in this research will provide some promising outcomes in depressivity prediction using EEG data. However, the highest correlation was obtained in the female-only group. In the future, we plan to increase the number of male samples to further explore the characteristics of the male group.

A key aspect of the current study is its prediction of a trait that is spread through the population and has high values linked to mental disorder. Our success opens the way to applying these methods to other traits linked to other mental disorders, such as anxiety. Such future work could explore both sensitivity and specificity of the prediction of multiple disorders and, since the measures are linked to trait scores, the issue of comorbidity. This will be particularly important for comorbid depression with anxiety, which appears to be particularly disabling.

References

Abásolo, D., Hornero, R., Espino, P., Poza, J., Sánchez, C. I., and de la Rosa, R. (2005). Analysis of regularity in the EEG background activity of Alzheimer's disease patients with Approximate Entropy. *Clinical Neurophysiology*, *116*(8), 1826–1834.

Acharya, U. R., Sree, S. V., Chattopadhyay, S., Yu, W., and Ang, P. C. A. (2011). Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. *International journal of neural systems*, *21*(03), 199–211.

Acharya, U. R., Sree, S. V., Swapna, G., Martis, R. J., and Suri, J. S. (2013). Automated EEG analysis of epilepsy: a review. *Knowledge-Based Systems*, *45*, 147–165.

Acharya, U. R., Sudarshan, V. K., Adeli, H., Santhosh, J., Koh, J. E., and Adeli, A. (2015). Computer-aided diagnosis of depression using EEG signals. *European neurology*, *73*(5-6), 329–336.

Acharya, U. R., Sudarshan, V. K., Adeli, H., Santhosh, J., Koh, J. E., Puthankatti, S. D., and Adeli, A. (2015). A novel depression diagnosis index using nonlinear features in EEG signals. *European neurology*, *74*(1-2), 79–83.

Ackermann, P., Kohlschein, C., Bitsch, J. A., Wehrle, K., and Jeschke, S. (2016). EEG-based automatic emotion recognition: Feature extraction, selection and classification methods. In *e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on*, 1–6. IEEE.

Adeli, H., Ghoshdastidar, S., and Dadmehr, N. (2010). Automated EEG-based diagnosis of neurological disorders: Inventing the future of neurology. *Crc Press*.

Adeli, H. and Hung, S. L. (1995). *Machine learning: neural networks, genetic algorithms, and fuzzy systems*. John Wiley & Sons, Inc.

Adeli, H. and Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural Networks*, *22*(7), 1018–1024. cited By 167.

Adeli, H. and Park, H. S. (1998). *Neurocomputing for design automation*. CRC Press, Inc.

Ahammad, N., Fathima, T., and Joseph, P. (2014). Detection of epileptic seizure event and onset using EEG. *BioMed research international*, *2014*.

Ahmadlou, M. and Adeli, H. (2010). Enhanced probabilistic neural network with local decision circles: A robust classifier. *Integrated Computer-Aided Engineering*, *17*(3), 197–210. cited By 133.

Ahmadlou, M., Adeli, H., and Adeli, A. (2010). Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder. *Journal of Clinical Neurophysiology*, *27*(5), 328–333.

Ahmadlou, M., Adeli, H., and Adeli, A. (2012a). Fractality analysis of frontal brain in major depressive disorder. *International Journal of Psychophysiology*, *85*(2), 206–211. cited By 86.

Ahmadlou, M., Adeli, H., and Adeli, A. (2012b). Fuzzy synchronization likelihood-wavelet methodology for diagnosis of autism spectrum disorder. *Journal of neuroscience methods*, *211*(2), 203–209.

Ahmadlou, M., Adeli, H., and Adeli, A. (2013). Spatiotemporal analysis of relative convergence of EEGs reveals differences between brain dynamics of depressive women and men. *Clinical EEG and neuroscience*, *44*(3), 175–181.

Akdemir Akar, S., Kara, S., Agambayev, S., and Bilgic, V. (2015). Nonlinear analysis of EEGs of patients with major depression during different emotional states. *Comput Biol Med*, *67*, 49–60.

Al-Nafjan, A., Hosny, M., Al-Ohali, Y., and Al-Wabil, A. (2017). Review and classification of emotion recognition based on EEG brain-computer interface system research: A systematic review. *Applied Sciences*, *7*(12), 1239.

Albert, P. R. (2015). Why is depression more prevalent in women? *Journal of psychiatry & neuroscience: JPN*, *40*(4), 219.

Allen, J. J., Coan, J. A., and Nazarian, M. (2004). Issues and assumptions on the road from raw signals to metrics of frontal EEG asymmetry in emotion. *Biological psychology*, *67*(1-2), 183–218.

Allen, J. J., Urry, H. L., Hitt, S. K., and Coan, J. A. (2004). The stability of resting frontal electroencephalographic asymmetry in depression. *Psychophysiology*, *41*(2), 269–280.

Alzheimer's Association (2018). What is Alzheimer's? `https://www.alz.org/alzheimers-dementia/what-is-alzheimers`. [Online; accessed 21-Sep-2018].

American Psychiatric Association (2017). What is depression? `https://www.psychiatry.org/patients-families/depression/what-is-depression`. [Online; accessed 15-Aug-2018].

American Psychiatric Association *et al.* (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

American Psychiatric Association *et al.* (2016). Online assessment measures: The personality inventory for DSM-5 (PID-5)-adults.

Andina, D. and Pham, D. T. (2007). *Computational intelligence: For engineering and manufacturing*. Springer.

Armitage, R., Hoffmann, R., Trivedi, M., and Rush, A. J. (2000). Slow-wave activity in NREM sleep: sex and age effects in depressed outpatients and healthy controls. *Psychiatry research*, *95*(3), 201–213.

Armitage, R. and Hoffmann, R. F. (2001). Sleep EEG, depression and gender. *Sleep medicine reviews*, *5*(3), 237–246.

Arnaud, Delorme and Scott, Makeig (2018). Decomposing Data Using ICA. `https://sccn.ucsd.edu/wiki/Chapter_09:_Decomposing_Data_Using_ICA`. Online; accessed 14-Dec-2019.

Aydin, S., Arica, N., Ergul, E., and Tan, O. (2015). Classification of obsessive compulsive disorder by EEG complexity and hemispheric dependency measurements. *International journal of neural systems*, *25*(03), 1550010.

Bachmann, M., Kalev, K., Suhhova, A., Lass, J., and Hinrikus, H. (2015). Lempel Ziv complexity of EEG in depression. In *6th European Conference of the International Federation for Medical and Biological Engineering*, 58–61. Springer.

Bachmann, M., Lass, J., and Hinrikus, H. (2017). Single channel EEG analysis for detection of depression. *Biomedical Signal Processing and Control*, *31*, 391–397.

Baehr, E., Rosenfeld, J. P., Baehr, R., and Earnest, C. (1998). Comparison of two EEG asymmetry indices in depressed patients vs. normal controls. *International Journal of Psychophysiology*, *31*(1), 89–92.

Banoczi, W. R. (2005). How some drugs affect the electroencephalogram (EEG). *American journal of electroneurodiagnostic technology*, *45*(2), 118–129.

Barnsley, M. F. (2014). *Fractals everywhere*. Academic press.

Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., and Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, *118*(12), 2765–2773.

Beck, A., Ward, C., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561–571. cited By 23102.

Behnam, H., Sheikhani, A., Mohammadi, M. R., Noroozian, M., and Golabi, P. (2008). Abnormalities in Connectivity of Quantitative Electroencephalogram Background Activity in Autism Disorders especially in Left Hemisphere and Right Temporal. In *Tenth international conference on computer modeling and simulation (uksim 2008)*, 82–87.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, *7*(6), 1129–1159.

Belmaker, R. and Agam, G. (2008). Major depressive disorder. *New England Journal of Medicine*, *358*(1), 55–68. PMID: 18172175.

Belouchrani, A., Abed-Meraim, K., Cardoso, J., and Moulines, E. (1993). Second-order blind separation of temporally correlated sources. In *Proc. Int. Conf. Digital Signal Processing*, 346–351. Citeseer.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*, *45*(2), 434–444.

Bengio, Y. and Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, *5*(Sep), 1089–1105.

Berg, P. and Scherg, M. (1991). Dipole modelling of eye activity and its application to the removal of eye artefacts from the EEG and MEG. *Clinical Physics and Physiological Measurement*, *12*(A), 49.

Berger, H. (1932). über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *97*(1), 6–26.

Besthorn, C., Förstl, H., Geiger-Kabisch, C., Sattel, H., Gasser, T., and Schreiter-Gasser, U. (1994). EEG coherence in Alzheimer disease. *Clinical Neurophysiology*, *90*(3), 242–245.

Besthorn, C., Sattel, H., Geiger-Kabisch, C., Zerfass, R., and Förstl, H. (1995). Parameters of EEG dimensional complexity in Alzheimer's disease. *Electroencephalography and Clinical Neurophysiology*, *95*(2), 84–89.

Bhat, S., Acharya, U. R., Adeli, H., Bairy, G. M., and Adeli, A. (2014). Automated diagnosis of autism: in search of a mathematical marker. *Reviews in the Neurosciences*, *25*(6), 851–861.

Bhatia, S. K. and Bhatia, S. C. (2007). Childhood and adolescent depression. *Depression*, *100*, 53.

Bianchi, L. (1895). The functions of the frontal lobes. *Brain*, *18*(4), 497–522.

Blackhart, G. C., Minnix, J. A., and Kline, J. P. (2006). Can EEG asymmetry patterns predict future development of anxiety and depression?: A preliminary study. *Biological psychology*, *72*(1), 46–50.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal processing magazine*, *25*(1), 41–56.

Blume, W. T. (2006). Drug effects on EEG. *Journal of Clinical Neurophysiology*, *23*(4), 306–311.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Breiman, L. and Cutler, A. (2019). Random Forests. `https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm`. Online; accessed 11-June-2019.

Brenner, R. P., Ulrich, R. F., Spiker, D. G., Sclabassi, R. J., Reynolds, C. F., Marin, R. S., and Boller, F. (1986). Computerized EEG spectral analysis in elderly normal, demented and depressed subjects. *Electroencephalography and Clinical Neurophysiology*, *64*(6), 483 – 492.

Britton, J., Frey, L. C., Hopp, J., Korb, P., Koubeissi, M., Lievens, W., Pestana-Knight, E., and St Louis, E. (2016). *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants.Appendix 4. Common Artifacts During EEG Recording*. American Epilepsy Society, Chicago.

Bruder, G., Fong, R., Tenke, C., Leite, P., Towey, J., Stewart, J., McGrath, P., and Quitkin, F. (1997). Regional brain asymmetries in major depression with or without an anxiety disorder: A quantitative electroencephalographic study. *Biological Psychiatry*, *41*(9), 939–948. cited By 226.

Bruder, G., Stewart, J., Tenke, C., McGrath, P., Leite, P., Bhattacharya, N., and Quitkin, F. (2001). Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an SSRI antidepressant. *Biological Psychiatry*, *49*(5), 416–425. cited By 119.

Brunner, C., Delorme, A., and Makeig, S. (2013). Eeglab–an open source matlab toolbox for electrophysiological research. *Biomed Tech*, *58*, 1.

Bryden, M. (1982). Laterality: Functional asymmetry in the intact. *Brain. Academic Press, New York*.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167. cited By 10393.

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *science*, *304*(5679), 1926–1929.

Camastra, F. and Vinciarelli, A. (2001). Cursive character recognition by learning vector quantization. *Pattern Recognition Letters*, *22*(6), 625–629.

Candy, J. V. (2016). *Bayesian signal processing: classical, modern, and particle filtering methods*, Volume 54. John Wiley & Sons.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proceedings of the IEEE*, *86*(10), 2009–2025.

Carver, C. S. and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of personality and social psychology*, *67*(2), 319.

Chua, C. K., Chandran, V., Acharya, R. U., and Min, L. C. (2009). Cardiac health diagnosis using higher order spectra and support vector machine. *Open Medical Informatics Journal*, *3*(1), 1–8.

Chua, K. C., Chandran, V., Acharya, R., and Lim, C. M. (2008a). Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: A comparative study. In *Engineering in Medicine and Biology Society, 2008. Embs 2008. International Conference of the IEEE*, 3824.

Chua, K. C., Chandran, V., Acharya, U. R., and Lim, C. M. (2008b). Cardiac state diagnosis using higher order spectra of heart rate variability. *Journal of Medical Engineering & Technology*, *32*(2), 145–155.

Chua, K. C., Chandran, V., Acharya, U. R., and Lim, C. M. (2009). Automatic identification of epileptic electroencephalography signals using higher-order spectra. *Proceedings of the Institution of Mechanical Engineers Part H Journal of Engineering in Medicine*, *223*(4), 485.

Cohen, L., Dehaene, S., Chochon, F., Lehericy, S., and Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia*, *38*(10), 1426–1440.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*(3), 287–314.

Congedo, M., Gouy-Pailler, C., and Jutten, C. (2008). On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clinical Neurophysiology*, *119*(12), 2677–2686.

Conn, R. (1946). The influence of emotion of the human EEG. *J. Nerv. Ment. Dis*, *104*, 351.

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, *19*(90), 297–301.

Croft, R. J. and Barry, R. J. (2000a). EOG correction: Which regression should we use? *Psychophysiology*, *37*(1), 123–125.

Croft, R. J. and Barry, R. J. (2000b). Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*(1), 5–19.

Croft, R. J. and Barry, R. J. (2000c). Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*(1), 5–19.

Daly, I., Nicolaou, N., Nasuto, S. J., and Warwick, K. (2013). Automated artifact removal from the electroencephalogram: a comparative study. *Clinical EEG and neuroscience*, *44*(4), 291–306.

Dangel, S., Meier, P., Moser, H., Plibersek, S., and Shen, Y. (1999). Time series analysis of sleep EEG. *Computer Assisted Physics*, *14*, 93–95. cited By 22.

Dawson, G., Frey, K., Panagiotides, H., Osterling, J., and Hessl, D. (1997). Infants of depressed mothers exhibit atypical frontal brain activity: A replication and extension of previous findings. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*(2), 179–186. cited By 105.

Debener, S., Beauducel, A., Nessler, D., Brocke, B., Heilemann, H., and Kayser, J. (2000). Is resting anterior EEG alpha asymmetry a trait marker for depression? *Neuropsychobiology*, *41*(1), 31–37.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, *134*(1), 9–21.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PloS one*, *7*(2), e30135.

Deslandes, A. C., de Moraes, H., Pompeu, F. A., Ribeiro, P., Cagy, M., Capitão, C., Alves, H., Piedade, R. A., and Laks, J. (2008). Electroencephalographic frontal asymmetry and depressive symptoms in the elderly. *Biological psychology*, *79*(3), 317–322.

Dhar, P., Dhar, A., Dey, S. K., Pal, D., Saha, S., and Goswami, P. (2017). EEG Signal classification by using Empirical Mode Decomposition and LVQ.

Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic bulletin & review*, *11*(6), 1011–1026.

Dietsch, G. (1932). Fourier-analyse von elektrencephalogrammen des menschen. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, *230*(1), 106–112.

Ekstrom, A. D., Caplan, J. B., Ho, E., Shattuck, K., Fried, I., and Kahana, M. J. (2005). Human hippocampal theta activity during virtual navigation. *Hippocampus*, *15*(7), 881–889.

Elbert, T., Lutzenberger, W., Rockstroh, B., Berg, P., and Cohen, R. (1992). Physical aspects of the EEG in schizophrenics. *Biological psychiatry*, *32*(7), 595–606.

Evinger, C., Shaw, M., Peck, C., Manning, K., and Baker, R. (1984). Blinking and associated eye movements in humans, guinea pigs, and rabbits. *Journal of Neurophysiology*, *52*(2), 323–339.

Eysenck, H. J. (1991). Manual of the Eysenck personality scales (EPS Adult). London: Hodder & Stoughton.

Fathima, T., Bedeeuzzaman, M., Farooq, O., and Khan, Y. (2011). Wavelet Based Features for Epileptic Seizure Detection. *MES Journal of Technology and Management*, *2(1)*.

Fatourechi, M., Bashashati, A., Ward, R. K., and Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clin Neurophysiol*, *118*(3), 480–94.

Faust, O., Acharya, U. R., Adeli, H., and Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure*, *26*, 56–64.

Faust, O., Ang, P. C. A., Puthankattil, S. D., and Joseph, P. K. (2014). Depression diagnosis support system based on EEG signal entropies. *Journal of mechanics in medicine and biology*, *14*(03), 1450035.

Fell, J., Röschke, J., and Beckmann, P. (1993). Deterministic chaos and the first positive Lyapunov exponent: a nonlinear analysis of the human electroencephalogram during sleep. *Biological cybernetics*, *69*(2), 139–146.

Fernández, A., López-Ibor, M.-I., Turrero, A., Santos, J.-M., Morón, M.-D., Hornero, R., Gómez, C., Méndez, M. A., Ortiz, T., and López-Ibor, J. J. (2011). Lempel–Ziv complexity in schizophrenia: a MEG study. *Clinical Neurophysiology*, *122*(11), 2227–2235.

Fisher, R. A. (1992). Statistical methods for research workers. In *Breakthroughs in statistics*, 66–70. Springer.

Fonti, V. and Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*.

Fossati, A., Krueger, R. F., Markon, K. E., Borroni, S., and Maffei, C. (2013). Reliability and validity of the personality inventory for DSM-5 (PID-5): predicting DSM-IV personality disorders and psychopathy in community-dwelling Italian adults. *Assessment*, *20*(6), 689–708.

Freund, Y., Schapire, R., and Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, *14*(771-780), 1612.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119–139.

Frid, A. (2014). Differences in phase synchrony of brain regions between regular and dyslexic readers. In *2014 IEEE 28th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, 1–4. IEEE.

Frid, A. and Manevitz, L. M. (2018). Features and Machine Learning for Correlating and Classifying between Brain Areas and Dyslexia. *arXiv preprint arXiv:1812.10622*.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics New York.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, *33*(1), 1.

Friedman, J., Hastie, T., Tibshirani, R., *et al.* (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, *28*(2), 337–407.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual review of neuroscience*, *32*, 209–224.

Fröhlich, F. (2016). *Network neuroscience*. Academic Press.

Ge, S., Wang, R., and Yu, D. (2014). Classification of four-class motor imagery employing single-channel electroencephalography. *PloS one*, *9*(6), e98019.

Giacinto, G., Perdisci, R., Del Rio, M., and Roli, F. (2008). Intrusion detection in computer networks by a modular ensemble of one-class classifiers. *Information Fusion*, *9*(1), 69–82.

Gilbert, D. G., McClernon, F. J., Rabinovich, N. E., Sugai, C., Plath, L. C., Asgaard, G., Zuo, Y., Huggenvik, J., and Botros, N. (2004). Effects of quitting smoking on EEG activation and attention last for more than 31 days and are more severe with stress, dependence, DRD2 A1 allele, and depressive traits. *Nicotine & Tobacco Research*, *6*(2), 249–267.

Gillies, G. E. and McArthur, S. (2010). Estrogen actions in the brain and the basis for differential action in men and women: a case for sex-specific medicines. *Pharmacological reviews*, pr–109.

Goodman, W. K. (2009). Research on biomarkers for mental disorders. `https://www.nimh.nih.gov/funding/grant-writing-and-application-process/concept-clearances/2009/research-on-biomarkers-for-mental-disorders.shtml`.

Gordon, E., Palmer, D. M., and Cooper, N. (2010). EEG alpha asymmetry in schizophrenia, depression, PTSD, panic disorder, ADHD and conduct disorder. *Clinical EEG and neuroscience*, *41*(4), 178–183.

Gotlib, I. H. (1998). EEG alpha asymmetry, depression, and cognitive functioning. *Cognition & Emotion*, *12*(3), 449–478.

Gotman, J., Skuce, D. R., Thompson, C. J., Gloor, P., Ives, J. R., and Ray, W. F. (1973). Clinical applications of spectral analysis and extraction of features from electroencephalograms with slow waves in adult patients. *Clinical Neurophysiology*, *35*(3), 225–235.

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 44–53.

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, *55*(4), 468–484.

Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., and Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal of clinical psychiatry*, *76*(2), 155–162.

Grin-Yatsenko, V. A., Baas, I., Ponomarev, V. A., and Kropotov, J. D. (2009). EEG power spectra at early stages of depressive disorders. *Journal of Clinical Neurophysiology*, *26*(6), 401–406.

Guo, Z. and Wong, W. (2013). 2 - Fundamentals of artificial intelligence techniques for apparel management applications. In W. Wong, Z. Guo, and S. Leung (Eds.), *Optimizing Decision Making in the Apparel Supply Chain Using Artificial Intelligence (AI)*, Woodhead Publishing Series in Textiles, 13 – 40. Woodhead Publishing.

Haglund, M. M., Berger, M. S., Shamseldin, M., Lettich, E., and Ojemann, G. A. (1994). Cortical localization of temporal lobe language sites in patients with gliomas. *Neurosurgery*, *34*(4), 567–576.

Halbreich, U. and Lumley, L. A. (1993). The multiple interactional biological processes that might lead to depression and gender differences in its appearance. *Journal of Affective Disorders*, *29*(2-3), 159–173.

Halder, S., Bensch, M., Mellinger, J., Bogdan, M., Kubler, A., Birbaumer, N., and Rosenstiel, W. (2007). Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Comput Intell Neurosci*, 82069.

Han, J. and Kamber, M. (2005). *Datamining: Concepts and techniques*. Morgan Kaufmann. cited By 677.

Hazarika, N., Chen, J. Z., Tsoi, A. C., and Sergejew, A. (1997). Classification of EEG signals using the wavelet transform. In *Digital Signal Processing Proceedings, 1997. DSP 97., 1997 13th International Conference on*, Volume 1, 89–92. IEEE.

He, P., Wilson, G., and Russell, C. (2004). Removal of ocular artifacts from electroencephalogram by adaptive filtering. *Medical and biological engineering and computing*, *42*(3), 407–412.

Health Jade Team (2018). Human brain. `https://healthjade.com/human-brain/`. [Online; accessed 09-Dec-2018].

Healthdirect Australia (2017). Diagnosis of depression. `https://www.healthdirect.gov.au/diagnosis-of-depression`. [Online; accessed 29-Aug-2018].

Henriques, J. and Davidson, R. (1990). Regional brain electrical asymmetries discriminate between previously depressed and healthy control subjects. *Journal of Abnormal Psychology*, *99*(1), 22–31. cited By 436.

Henriques, J. B. and Davidson, R. J. (1991). Left frontal hypoactivation in depression. *Journal of abnormal psychology*, *100*(4), 535.

Hérault, J. and Ans, B. (1984). Neuronal network with modifiable synapses: decoding of composite sensory messages under unsupervised and permanent learning. *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie*, *299*(13), 525–528.

Hochreiter, S. and Schmidhuber, J. (1997). Low-complexity coding and decoding. *Theoretical aspects of neural computation (TANC 97), Hong Kong*, 297–306.

Hosseinifard, B., Moradi, M. H., and Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput Methods Programs Biomed*, *109*(3), 339–45.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417.

Hsu, W.-Y. (2013). SINGLE-trial motor imagery classification using asymmetry ratio, phase relation, wavelet-based fractal, and their selected combination. *International Journal of Neural Systems*, *23*(2). cited By 51.

Hu, X., Wang, B., and Ji, H. (2013). A wireless sensor network-based structural health monitoring system for highway bridges. *Computer-Aided Civil and Infrastructure Engineering*, *28*(3), 193–209.

Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 20110534.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural computation*, *9*(7), 1483–1492.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, *13*(4-5), 411–430.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders.

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry*, *171*(4), 395–397.

Inuso, G., La Foresta, F., Mammone, N., and Morabito, F. C. (2007). Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings. In *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, 1524–1529. IEEE.

Ithaya Rani, P. and Muneeswaran, K. (2016). Facial emotion recognition based on eye and mouth regions. *International Journal of Pattern Recognition and Artificial Intelligence*, *30*(07), 1655020.

Izzetoglu, M., Devaraj, A., Bunce, S., and Onaral, B. (2005). Motion artifact cancellation in NIR spectroscopy using Wiener filtering. *IEEE Transactions on Biomedical Engineering*, *52*(5), 934–938.

Jahankhani, P., Kodogiannis, V., and Revett, K. (2006). EEG signal classification using wavelet feature extraction and neural networks. In *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*, 120–124. IEEE.

James, C. J. and Hesse, C. W. (2005). Independent component analysis for biomedical signals. *Physiological measurement*, *26*(1), R15.

Jelles, B., Strijers, R., Hooijer, C., Jonker, C., Stam, C., and Jonkman, E. (1999). Nonlinear EEG analysis in early Alzheimer's disease. *Acta neurologica scandinavica*, *100*(6), 360–368.

Jeong, J. (2002). Nonlinear dynamics of EEG in Alzheimer's disease. *Drug Development Research*, *56*(2), 57–66.

Jetter, W., Poser, U., Freeman Jr, R. B., and Markowitsch, H. J. (1986). A verbal long term memory deficit in frontal lobe damaged patients. *Cortex*, *22*(2), 229–242.

Jiang, X. and Adeli, H. (2004). Wavelet packet-autocorrelation function method for traffic flow pattern analysis. *Computer-Aided Civil and Infrastructure Engineering*, *19*(5), 324–337.

Jiang, X., Mahadevan, S., and Adeli, H. (2007). Bayesian wavelet packet denoising for structural system identification. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures*, *14*(2), 333–356.

Jie, X., Cao, R., and Li, L. (2014). Emotion recognition based on the sample entropy of EEG. *Bio-medical materials and engineering*, *24*(1), 1185–1192.

Jung, T., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human brain mapping*, *14*(3), 166–185.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745–1758.

Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics*, *30*(1), 197–215.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35–45.

Kelsey, C. and G.Fischer, D. (2015). The PID-5 does not contain validity scales. Do you care?

Kemp, A., Griffiths, K., Felmingham, K., Shankman, S. A., Drinkenburg, W., Arns, M., Clark, C. R., and Bryant, R. (2010). Disorder specificity despite comorbidity:

resting EEG alpha asymmetry in major depressive disorder and post-traumatic stress disorder. *Biological Psychology*, *85*(2), 350–354.

Kessler, R. C., DuPont, R. L., Berglund, P., and Wittchen, H.-U. (1999). Impairment in pure and comorbid generalized anxiety disorder and major depression at 12 months in two national surveys. *American journal of psychiatry*, *156*(12), 1915–1923.

Kim, D.-J., Jeong, J., Chae, J.-H., Park, S., Kim, S. Y., Go, H. J., Paik, I.-H., Kim, K.-S., and Choi, B. (2000). An estimation of the first positive Lyapunov exponent of the EEG in patients with schizophrenia. *Psychiatry Research: Neuroimaging*, *98*(3), 177–189.

Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational statistics & data analysis*, *53*(11), 3735–3745.

Kim, M. and Kim, S.-P. (2018). A comparsion of artifact rejection methods for a BCI using event related potentials. In *Brain-Computer Interface (BCI), 2018 6th International Conference on*, 1–4. IEEE.

Kimberg, D. Y. and Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*(4), 411.

Kimura, D. (1961). Some effects of temporal-lobe damage on auditory perception. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *15*(3), 156.

Kirkove, M., François, C., and Verly, J. (2014). Comparative evaluation of existing and new methods for correcting ocular artifacts in electroencephalographic recordings. *Signal Processing*, *98*, 102–120.

Knott, V., Mahoney, C., Kennedy, S., and Evans, K. (2001). EEG power, frequency, asymmetry and coherence in male depression. *Psychiatry Research: Neuroimaging*, *106*(2), 123–140.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

Kohonen, T. (1995). *Learning Vector Quantization*, 175–189. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kohonen, T. and Maps, S.-O. (1995). Springer series in information sciences. *Self-organizing maps*, *30*.

Koles, Z., Lind, J., and Flor-Henry, P. (1994). Spatial patterns in the background EEG underlying mental disease in man. *Electroencephalography and clinical neurophysiology*, *91*(5), 319–328.

Koles, Z. J., Lazar, M. S., and Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain topography*, *2*(4), 275–284.

Krajewski, J., Batliner, A., and Kessel, S. (2010). Comparing multiple classifiers for speech-based detection of self-confidence-A pilot study. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, 3716–3719. IEEE.

Krishnaveni, V., Jayaraman, S., Kumar, P. M., Shivakumar, K., and Ramadoss, K. (2005). Comparison of independent component analysis algorithms for removal of ocular artifacts from electroencephalogram. *Measurement Science Review*, *5*(2), 67–78.

Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., and Skodol, A. E. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological medicine*, *42*(9), 1879–1890.

Krueger, R. F. and Markon, K. E. (2014). The role of the DSM-5 personality trait model in moving toward a quantitative and empirically based approach to classifying personality and psychopathology. *Annu Rev Clin Psychol*, *10*, 477–501.

Krystal, A. D., Zaidman, C., Greenside, H. S., Weiner, R. D., and Coffey, C. E. (1997). The largest Lyapunov exponent of the EEG during ECT seizures as a measure of ECT seizure adequacy. *Electroencephalography and clinical neurophysiology*, *103*(6), 599–606.

Kupfer, D. J., Frank, E., and Phillips, M. L. (2012). Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *The Lancet*, *379*(9820), 1045–1055.

Kuskowski, M. A., Malone, S. M., Kim, S. W., Dysken, M. W., Okaya, A. J., and Christensen, K. J. (1993). Quantitative EEG in obsessive-compulsive disorder. *Biological psychiatry*, *33*(6), 423–430.

Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, 556–562.

Lee, J.-M., Kim, D.-J., Kim, I.-Y., Park, K.-S., and Kim, S. (2002). Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data. *Computers in Biology and Medicine*, *32*(1), 37–47. cited By 83.

Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, *11*(2), 417–441.

Lega, B. C., Jacobs, J., and Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, *22*(4), 748–761.

Levy, W. J. (1987). Effect of epoch length on power spectrum analysis of the EEG. *Anesthesiology*, *66*(4), 489–495.

Li, M. and Lu, B.-L. (2009). Emotion classification based on gamma-band EEG. In *Engineering in medicine and biology society, 2009. EMBC 2009. Annual international conference of the IEEE*, 1223–1226. IEEE.

Li, Y., Tong, S., Liu, D., Gai, Y., Wang, X., Wang, J., Qiu, Y., and Zhu, Y. (2008). Abnormal EEG complexity in patients with schizophrenia and depression. *Clin Neurophysiol*, *119*(6), 1232–41.

Li, Y.-j. and Fan, F.-y. (2006). Classification of schizophrenia and depression by EEG with ANNs. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2679–2682. IEEE.

Liao, S.-C., Wu, C.-T., Huang, H.-C., Cheng, W.-T., and Liu, Y.-H. (2017). Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns. *Sensors*, *17*(6), 1385.

Lindgren, K. A., Larson, C. L., Schaefer, S. M., Abercrombie, H. C., Ward, R. T., Oakes, T. R., Holden, J. E., Perlman, S. B., Benca, R. M., and Davidson, R. J. (1999). Thalamic metabolic rate predicts EEG alpha power in healthy control subjects but not in depressed patients. *Biological psychiatry*, *45*(8), 943–952.

Locasto, M. E., Wang, K., Keromytis, A. D., and Stolfo, S. J. (2005). Flips: Hybrid adaptive intrusion prevention. In *International Workshop on Recent Advances in Intrusion Detection*, 82–101. Springer.

Loh, W.-Y. and Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica sinica*, 815–840.

Luo, D., Ibrahim, Z., Ismail, Z., and Xu, B. (2013). Optimization of the geometries of biconical tapered fiber sensors for monitoring the early-age curing temperatures of concrete specimens. *Computer-Aided Civil and Infrastructure Engineering*, *28*(7), 531–541.

Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In *Advances in neural information processing systems*, 145–151.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, *11*(7), 674–693.

Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *science*, *156*(3775), 636–638.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*, Volume 1. WH freeman New York.

Markon, K. E., Quilty, L. C., Bagby, R. M., and Krueger, R. F. (2013). The development and psychometric properties of an informant-report form of the personality inventory for DSM-5 (PID-5). *Assessment*, *20*(3), 370–83.

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5), 237–329.

Marwan, N., Thiel, M., and Nowaczyk, N. R. (2002). Cross recurrence plot based synchronization of time series. *Nonlinear Processes in Geophysics,9,3/4(-0001-11-30)*, *9*(3), 325–331.

Matiko, J. W., Beeby, S., and Tudor, J. (2013). Real time eye blink noise removal from EEG signals using morphological component analysis. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 13–16. IEEE.

McFarland, D. J., McCane, L. M., David, S. V., and Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalography and clinical Neurophysiology*, *103*(3), 386–394.

Merkl, A., Neumann, W.-J., Huebl, J., Aust, S., Horn, A., Krauss, J. K., Dziobek, I., Kuhn, J., Schneider, G.-H., Bajbouj, M., *et al.* (2015). Modulation of beta-band activity in the subgenual anterior cingulate cortex during emotional empathy in treatment-resistant depression. *Cerebral Cortex*, *26*(6), 2626–2638.

Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–40.

Mohammadi, M., Al-Azab, F., Raahemi, B., Richards, G., Jaworska, N., Smith, D., de la Salle, S., Blier, P., and Knott, V. (2015). Data mining EEG signals in depression for their diagnostic value. *BMC medical informatics and decision making*, *15*(1), 108.

Mu, X., Lu, J., Watta, P., and Hassoun, M. H. (2009). Weighted voting-based ensemble classifiers with application to human face recognition and voice recognition. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, 2168–2171. IEEE.

Müller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical neurophysiology*, *110*(5), 787–798.

Mumtaz, W., Xia, L., Yasin, M. A. M., Ali, S. S. A., and Malik, A. S. (2017). A wavelet-based technique to predict treatment outcome for major depressive disorder. *PloS one*, *12*(2), e0171409.

Mustafa, M., Taib, M., Murat, Z., and Sulaiman, N. (2012). Comparison between KNN and ANN classification in brain balancing application via spectrogram image. *Journal of Computer Science & Computational Mathematics*, *2*(4), 17–22.

Muthuswamy, J., Sherman, D. L., and Thakor, N. V. (1999). Higher-order spectral analysis of burst patterns in EEG. *IEEE Transactions on Biomedical Engineering*, *46*(1), 92–99.

Nabbe, P., Le Reste, J. Y., Guillou-Landreat, M., Munoz Perez, M. A., Argyriadou, S., Claveria, A., Fernandez San Martin, M. I., Czachowski, S., Lingner, H., Lygidakis, C., Sowinska, A., Chiron, B., Derriennic, J., Le Prielec, A., Le Floch, B., Montier, T., Van Marwijk, H., and Van Royen, P. (2017). Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review. *Eur Psychiatry*, *39*, 99–105.

National Institute of Mental Health (2018a). Depression. https://www.nimh.nih.gov/health/topics/depression/index.shtml. [Online; accessed 15-Aug-2018].

National Institute of Mental Health (2018b). Depression: What you need to know. https://www.nimh.nih.gov/health/publications/depression-what-you-need-to-know/index.shtml. [Online; accessed 09-Dec-2018].

National Instruments (2019). Understanding FFTs and Windowing. https://www.ni.com/en-nz/innovations/white-papers/06/understanding-ffts-and-windowing.html. Online; accessed 13-Dec-2019.

New Zealand Parliamentary Counsel Office (2010). Electricity (Safety) Regulations 2010. http://www.legislation.govt.nz/regulation/public/2010/0036/latest/whole.html#DLM2763652. Online; accessed 15-May-2019.

Nofzinger, E. A., Price, J. C., Meltzer, C. C., Buysse, D. J., Villemagne, V. L., Miewald, J. M., Sembrat, R. C., Steppe, D. A., and Kupfer, D. J. (2000). Towards a neurobiology of dysfunctional arousal in depression: the relationship between beta EEG power and regional cerebral glucose metabolism during NREM sleep. *Psychiatry Research: Neuroimaging*, *98*(2), 71–91.

Nolan, H., Whelan, R., and Reilly, R. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of neuroscience methods*, *192*(1), 152–162.

Ocak, H. (2008). Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal processing*, *88*(7), 1858–1867.

Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, *36*(2), 2027–2036.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572.

Piccinelli, M. and Wilkinson, G. (2000). Gender differences in depression. *The British Journal of Psychiatry*, *177*(6), 486–492.

Pijn, J. P. M., Velis, D. N., van der Heyden, M. J., DeGoede, J., van Veelen, C. W., and da Silva, F. H. L. (1997). Nonlinear dynamics of epileptic seizures on basis of intracranial EEG recordings. *Brain topography*, *9*(4), 249–270.

Pincus, S. (1995). Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *5*(1), 110–117.

Pincus, S. and Keefe, D. (1992). Quantification of hormone pulsatility via an approximate entropy algorithm. *American Journal of Physiology - Endocrinology and Metabolism*, *262*(5 25-5), E741–E754. cited By 234.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, *88*(6), 2297–2301.

Pontifex, M. B., Gwizdala, K. L., Parks, A. C., Billinger, M., and Brunner, C. (2017). Variability of ICA decomposition may impact EEG signals when used to remove eyeblink artifacts. *Psychophysiology*, *54*(3), 386–398.

Poulos, M., Rangoussi, M., and Alexandris, N. (1999). Neural network based person identification using EEG features. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, Volume 2, 1117–1120. IEEE.

Poulos, M., Rangoussi, M., Alexandris, N., Evangelou, A., *et al.* (2002). Person identification from the EEG using nonlinear signal classification. *Methods of information in Medicine*, *41*(1), 64–75.

Pradhan, C., Jena, S. K., Nadar, S. R., and Pradhan, N. (2012). Higher-order spectrum in understanding nonlinearity in EEG rhythms. *Computational and mathematical methods in medicine*, *2012*.

Pradhan, N. and Dutt, D. N. (1993). Use of running fractal dimension for the analysis of changing patterns in electroencephalograms. *Computers in biology and medicine*, *23*(5), 381–388.

Pradhan, N., Sadasivan, P., and Arunodaya, G. (1996). Detection of seizure activity in EEG by an artificial neural network: A preliminary study. *Computers and Biomedical Research*, *29*(4), 303–313.

Pritchard, W. S., Duke, D. W., Coburn, K. L., Moore, N. C., Tucker, K. A., Jann, M. W., and Hostetler, R. M. (1994). EEG-based, neural-net predictive classifica-

tion of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures. *Electroencephalography and clinical Neurophysiology*, *91*(2), 118–130.

Provins, K. and Cunliffe, P. (1972). The relationship between EEG activity and handedness. *Cortex*, *8*(2), 136–146.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W., LaMantia, A., McNamara, J., and Williams, S. (2004). Neuroscience, 3rd edn. *Sunderland, MA: Sinauer Associates*.

Puthankattil, S. D. and Joseph, P. K. (2012a). Classification of EEG signals in normal and depression condition by ANN using RWE and signal entropy. *Journal of Mechanics in Medicine & Biology*, *12*(04), 528–1437.

Puthankattil, S. D. and Joseph, P. K. (2012b). Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy. *Journal of Mechanics in Medicine and Biology*, *12*(04), 1240019.

Puthankattil, S. D. and Joseph, P. K. (2014). Analysis of EEG signals using wavelet entropy and approximate entropy: A case study on depression patients. *Int. J. Med. Health Biomed. Pharm. Eng*, *8*(7), 420–424.

Raghavendra, B. and Dutt, D. N. (2010). Computing fractal dimension of signals using multiresolution box-counting method. *International Journal of Information and Mathematical Sciences*, *6*(1), 50–65.

Rani, P. I. and Muneeswaran, K. (2017). Recognize the facial emotion in video sequences using eye and mouth temporal Gabor features. *Multimedia Tools and Applications*, *76*(7), 10017–10040.

Rani, P. I. and Muneeswaran, K. (2018). Emotion recognition based on facial components. *Sādhanā*, *43*(3), 48.

Reddy, A. G. and Narava, S. (2013). Artifact removal from EEG signals. *International Journal of Computer Applications*, *77*(13), 17–19.

Reid, S., Duke, L., and Allen, J. (1998). Resting frontal electroencephalographic asymmetry in depression: Inconsistencies suggest the need to identify mediating factors. *Psychophysiology*, *35*(4), 389–404. cited By 246.

Reynolds III, C. F., Kupfer, D. J., Thase, M. E., Frank, E., Jarrett, D. B., Coble, P. A., Hoch, C. C., Buysse, D. J., Simons, A. D., and Houck, P. R. (1990). Sleep,

gender, and depression: an analysis of gender effects on the electroencephalographic sleep of 302 depressed outpatients. *Biological Psychiatry*, *28*(8), 673–684.

Richman, J. and Moorman, J. (2000). Physiological time-series analysis using approximate and sample entropy. *American Journal of Physiology - Heart and Circulatory Physiology*, *278*(6 47-6), H2039–H2049. cited By 2819.

Rokach, L. and Maimon, O. Z. (2008). *Data mining with decision trees: theory and applications*, Volume 69. World scientific.

Rolls, E. T. (2001). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. In *Vision: The Approach of Biophysics and Neurosciences*, 366–395. World Scientific.

Romero, L. M. R. (2013). *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013: MEDICON 2013, 25-28 September 2013, Seville, Spain*, Volume 41. Springer Science & Business Media.

Röschke, J. and Aldenhoff, J. (1991). The dimensionality of human's electroencephalogram during sleep. *Biological Cybernetics*, *64*(4), 307–313.

Rösler, O. and Suendermann, D. (2013). A first step towards eye state prediction using EEG. *Proc. of the AIHLS*.

Sahonero, G. and Calderon, H. (2017). A Comparison of SOBI, FastICA, JADE and Infomax Algorithms.

Saletu, B., Grünberger, J., Rajna, P., and Karobath, M. (1980). Clovoxamine and fluvoxamine-2 biogenic amine re-uptake inhibiting antidepressants: quantitative EEG, psychometric and pharmacokinetic studies in man. *Journal of neural transmission*, *49*(1-2), 63–86.

Salkind, N. J. (2010). *Encyclopedia of research design*, Volume 1. Sage.

Sankari, Z. and Adeli, H. (2011). Probabilistic neural networks for diagnosis of Alzheimer's disease using conventional and wavelet coherence. *Journal of Neuroscience Methods*, *197*(1), 165–170. cited By 107.

Sauseng, P., Hoppe, J., Klimesch, W., Gerloff, C., and Hummel, F. C. (2007). Dissociation of sustained attention from central executive functions: local activity and inter-

regional connectivity in the theta range. *European Journal of Neuroscience*, *25*(2), 587–593.

Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine Learning*, *13*(1), 135–143.

Schaffer, C., Davidson, R., and Saron, C. (1983). Frontal and parietal electroencephalogram asymmetry in depressed and nondepressed subjects. *Biological Psychiatry*, *18*(7), 753–762. cited By 257.

Schlogl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., and Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clin Neurophysiol*, *118*(1), 98–104.

Schmidt, H. D., Shelton, R. C., and Duman, R. S. (2011). Functional biomarkers of depression: diagnosis, treatment, and pathophysiology. *Neuropsychopharmacology*, *36*(12), 2375.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., and Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of neurophysiology*, *79*(6), 3272–3278.

Segrave, R. A., Cooper, N., Thomson, R., Croft, R. J., Sheppard, D., and Fitzgerald, P. (2011). Individualized alpha activity and frontal asymmetry in major depression. *Clinical EEG and Neuroscience*, *42*(1), 45–52.

Seni, G. and Elder, J. F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis lectures on data mining and knowledge discovery*, *2*(1), 1–126.

Shabtai, A., Moskovitch, R., Elovici, Y., and Glezer, C. (2009). Detection of malicious code by applying machine learning classifiers on static features: A state-of-the-art survey. *information security technical report*, *14*(1), 16–29.

Shackman, A. J., McMenamin, B. W., Maxwell, J. S., Greischar, L. L., and Davidson, R. J. (2009). Right dorsolateral prefrontal cortical activity and behavioral inhibition. *Psychological science*, *20*(12), 1500–1506.

Sheikhani, A., Behnam, H., Mohammadi, M., and Noroozian, M. (2007). Analysis of EEG background activity in Autism disease patients with bispectrum and STFT

measure. In *Proceedings of the 11th WSEAS International Conference on COMMUNICATIONS*.

Sheikhani, A., Behnam, H., Mohammadi, M. R., Noroozian, M., and Golabi, P. (2008). Connectivity analysis of quantitative Electroencephalogram background activity in Autism disorders with short time Fourier transform and Coherence values. In *2008 Congress on Image and Signal Processing*, Volume 1, 207–212. IEEE.

Shuttleworth, M. (2009). Counterbalanced measures design. `https://explorable.com/counterbalanced-measures-design`. [Online; accessed 20-Dec-2018].

Siddique, N. and Adeli, H. (2013). *Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing*. John Wiley & Sons.

Siuly (2012). *Analysis and classification of EEG signals*. Ph. D. thesis, University of Southern Queensland.

Smit, D., Posthuma, D., Boomsma, D., and De Geus, E. (2007). The relation between frontal EEG asymmetry and the risk for anxiety and depression. *Biological psychology*, *74*(1), 26–33.

Song, I.-H., Lee, D.-S., and Kim, S. I. (2004). Recurrence quantification analysis of sleep electoencephalogram in sleep apnea syndrome in humans. *Neuroscience letters*, *366*(2), 148–153.

Song, Y., Crowcroft, J., and Zhang, J. (2012). Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. *Journal of neuroscience methods*, *210*(2), 132–146.

Spielberger, C. D. and Gorsuch, R. (1983). *State-trait anxiety inventory (form Y)*. Consulting Psychologists Press.

Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberger, A., and Gaul, W. A. (2006). *From data and information analysis to knowledge engineering: proceedings of the 29th Annual Conference of the Gesellschaft für Klassifikation eV, University of Magdeburg, March 9-11, 2005*. Springer Science & Business Media.

Squire, L. R. and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Srinivasan, V., Eswaran, C., and Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Transactions on information Technology in Biomedicine*, *11*(3), 288–295.

Stewart, J. L. and Allen, J. J. (2018). Resting frontal brain asymmetry is linked to future depressive symptoms in women. *Biological Psychology*, *136*, 161–167.

Stuss, D. T. and Benson, D. F. (1986). *The frontal lobes*. Raven Pr.

Stuss, D. T., Gow, C. A., and Hetherington, C. R. (1992). " No longer gage": frontal lobe dysfunction and emotional changes. *Journal of consulting and clinical psychology*, *60*(3), 349.

Su, W., Huang, C., Chen, C., Liu, C., Huang, H., and Le, Q. (2014). Identifying the modal parameters of a structure from ambient vibration data via the stationary wavelet packet. *Computer-Aided Civil and Infrastructure Engineering*, *29*(10), 738–757.

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, *32*(4), 1084–1093.

Subha, D. P., Joseph, P. K., Acharya, R., and Lim, C. M. (2010). EEG signal analysis: a survey. *Journal of medical systems*, *34*(2), 195–212.

Subhash Chandran, K., Mishra, A., Shirhatti, V., and Ray, S. (2016). Comparison of matching pursuit algorithm with other signal processing techniques for computation of the time-frequency power spectrum of brain signals. *The Journal of Neuroscience*, *36*(12), 3399–3408.

Suman, J. V., Sumabindu, Y., and Seventline, J. B. (2015). Performance analysis of time frequency resolution techniques for non-stationary signals. *Indian Journal of Science and Technology*, *8*(23).

Sweeney, K. T., Ward, T. E., and McLoone, S. F. (2012). Artifact removal in physiological signals — practices and possibilities. *IEEE transactions on information technology in biomedicine*, *16*(3), 488–500.

Tamura, H. and Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, *11*(5), 384–399.

Teitti, S., Määttä, S., Säisänen, L., Könönen, M., Vanninen, R., Hannula, H., Mervaala, E., and Karhu, J. (2008). Non-primary motor areas in the human frontal lobe are connected directly to hand muscles. *Neuroimage*, *40*(3), 1243–1250.

Thau, K., Rappelsberger, P., Lovrek, A., Petsche, H., Simhandl, C., and Topitz, A. (1988). Effect of lithium on the EEG of healthy males and females: A probability mapping study. *Neuropsychobiology*, *20*(3), 158–163. cited By 21.

The MathWorks, Inc. (2018a). Correlation coefficients. `https://explorable.com/counterbalanced-measures-design`. Online; accessed 01-Jan-2019.

The MathWorks, Inc. (2018b). Predictor importance. `https://au.mathworks.com/help/stats/compactregressionensemble.predictorimportance.html`. [Online; accessed 21-Sep-2018].

The MathWorks, Inc. (2018c). Treebagger class. `https://au.mathworks.com/help/stats/treebagger-class.html#bvghisl`. [Online; accessed 14-Dec-2018].

Thibodeau, R., Jorgensen, R. S., and Kim, S. (2006). Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review. *Journal of abnormal psychology*, *115*(4), 715.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time–frequency analysis. *IEEE transactions on information technology in biomedicine*, *13*(5), 703–710.

Urigüen, J. A. and Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, *12*(3), 031001.

Vapnik, V. (1998). Statistical Learning Theory. *Wiley, New York*. cited By 19774.

Verleger, R. (1991). The instruction to refrain from blinking affects auditory P3 and N1 amplitudes. *Clinical Neurophysiology*, *78*(3), 240–251.

Verleger, R. (1993). Valid identification of blink artefacts: are they larger than 50 μV in EEG records? *Electroencephalography and clinical Neurophysiology*, *87*(6), 354–363.

Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and clinical neurophysiology*, *103*(3), 395–404.

Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and clinical neurophysiology*, *103*(3), 395–404.

Voets, N., Adcock, J., Flitney, D., Behrens, T., Hart, Y., Stacey, R., Carpenter, K., and Matthews, P. (2005). Distinct right frontal lobe activation in language processing following left hemisphere injury. *Brain*, *129*(3), 754–766.

Volkert, J., Schulz, H., Härter, M., Wlodarczyk, O., and Andreas, S. (2013). The prevalence of mental disorders in older people in Western countries–a meta-analysis. *Ageing research reviews*, *12*(1), 339–353.

Vuckovic, A., Radivojevic, V., Chen, A. C., and Popovic, D. (2002). Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Medical engineering & physics*, *24*(5), 349–360.

Wagner, A. D., Shannon, B. J., Kahn, I., and Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in cognitive sciences*, *9*(9), 445–453.

Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., and Sweeney, J. A. (2013). Resting state EEG abnormalities in autism spectrum disorders. *Journal of neurodevelopmental disorders*, *5*(1), 24.

Wang, Y., Gao, S., and Gao, X. (2006). Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 5392–5395. IEEE.

Wang, Y., Zhou, W., Yuan, Q., Li, X., Meng, Q., Zhao, X., and Wang, J. (2013). Comparison of ictal and interictal EEG signals using fractal features. *International Journal of Neural Systems*, *23*(6). cited By 26.

Webb, C. A., Dillon, D. G., Pechtel, P., Goer, F. K., Murray, L., Huys, Q. J., Fava, M., McGrath, P. J., Weissman, M., Parsey, R., *et al.* (2016). Neural correlates of

three promising endophenotypes of depression: evidence from the EMBARC study. *Neuropsychopharmacology*, *41*(2), 454.

Winkler, I., Debener, S., Müller, K.-R., and Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 4101–4105. IEEE.

Witelson, S. F. (1985). The brain connection: the corpus callosum is larger in left-handers. *Science*, *229*(4714), 665–668.

Woestenburg, J., Verbaten, M., and Slangen, J. (1983). The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. *Biological psychology*, *16*(1-2), 127–147.

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, *16*(3), 285–317.

World Health Organization (2017). Depression: let's talk. `http://www.who.int/mental_health/management/depression/en/`. [Online; accessed 09-Dec-2018].

Woyshville, M. J. and Calabrese, J. R. (1994). Quantification of occipital EEG changes in Alzheimer's disease utilizing a new metric: the fractal dimension. *Biological psychiatry*, *35*(6), 381–387.

Yong, X., Ward, R. K., and Birch, G. E. (2009). Artifact removal in EEG using morphological component analysis. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 345–348. IEEE.

Zabelina, D. L., Condon, D., and Beeman, M. (2014). Do dimensional psychopathology measures relate to creative achievement or divergent thinking? *Frontiers in psychology*, *5*, 1029.

Zeng, H., Song, A., Yan, R., and Qin, H. (2013). EOG artifact correction from EEG recording using stationary subspace analysis and empirical mode decomposition. *Sensors*, *13*(11), 14839–14859.

Zhang, B., Yin, J., Hao, J., Zhang, D., and Wang, S. (2007). Malicious codes detection based on ensemble learning. In *International Conference on Autonomic and Trusted Computing*, 468–477. Springer.

Zhang, C., Wang, H., and Wu, M.-H. (2013). EEG-based expert system using complexity measures and probability density function control in alpha sub-band. *Integrated Computer-Aided Engineering*, *20*(4), 391–405. cited By 28.

Zhang, S., McIntosh, J., Shadli, S. M., Neo, P. S., Huang, Z., and McNaughton, N. (2017). Removing eye blink artefacts from EEG-A single-channel physiology-based method. *J Neurosci Methods*, *291*, 213–220.